



# Application de l'apprentissage artificiel à la prévision des crues éclair

Mohamed Samir Toukourou

## ► To cite this version:

Mohamed Samir Toukourou. Application de l'apprentissage artificiel à la prévision des crues éclair. Hydrologie. École Nationale Supérieure des Mines de Paris, 2009. Français. NNT : 2009ENMP1669 . pastel-00005626

**HAL Id: pastel-00005626**

**<https://pastel.archives-ouvertes.fr/pastel-00005626>**

Submitted on 7 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ED n°431. : INFORMATION, COMMUNICATION, MODELISATION ET SIMULATION

# **T H E S E**

pour obtenir le grade de

**DOCTEUR DE L'ECOLE NATIONALE SUPERIEURE DES MINES DE PARIS**

Spécialité “Informatique Temps Réel, Robotique et Automatique”

présentée et soutenue publiquement par  
**Mohamed Samir TOUKOUROU**

le 10 décembre 2009

<b>Application de l'apprentissage artificiel à la prévision des crues éclair</b>
--

*Directeur de thèse : M. Gérard DREYFUS*

## **Jury**

M. D. DARTUS	Professeur	Rapporteur
M. C. TOUZET	Maître de Conférences	Rapporteur
M. G. DREYFUS	Professeur	Directeur de Thèse
Mme A. JOHANNET	Chargée de Recherches	Maître de Thèse
M. G.M. SAULNIER	Chargé de Recherches	Examineur
M. Y. VIMONT	Professeur	Examineur
M. B. JANET	SCHAPI	Invité



## Remerciements

J'aimerais adresser mes plus vifs remerciements à :

*Monsieur Gérard DREYFUS*, Professeur à l'Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI – ParisTech) et Directeur du Laboratoire d'Electronique de cet établissement, dont l'expertise sur le traitement des informations par les systèmes neuronaux a permis de mener cette recherche à bon port.

*Madame Anne JOHANNET*, Chargée de Recherches dans le pôle GéoMécanique & Hydrologie (GMH) du Centre des Matériaux de Grande Diffusion (CMGD) de l'Ecole des Mines d'Alès (EMA), qui fut mon encadrante de proximité. Elle a mené avec rigueur et disponibilité le déroulement de cette thèse. Je tiens à la remercier, aussi bien pour la confiance qu'elle m'a témoignée en m'accordant cette chance de réaliser ce projet de recherche, que pour sa participation active dans la validation des plans d'expérience menés tout au long de ces trois années.

*Madame Mireille FOULETIER* et *Monsieur Yannick VIMONT*, qui se sont succédés au poste de Directeur de mon laboratoire de rattachement, à savoir le CMGD. Je remercie la première pour avoir accepté mon recrutement en thèse dans son laboratoire et le second pour sa disponibilité et ses remarques pertinentes lors des différentes présentations durant lesquelles il a manifesté un grand intérêt pour mes travaux.

*Monsieur Yves ROUCHALEAU* et *Monsieur Jean-Paul MARMORAT*, respectivement Professeur et Maître de Recherche du Centre de Mathématiques Appliquées (CMA) de l'Ecole Nationale Supérieure des Mines de Paris (ENSM - ParisTech), qui m'ont accueilli dans leur spécialité doctorale à savoir Informatique Temps Réel, Robotique Automatique (IT2RA).

*Monsieur François GOULETTE*, Enseignant-Chercheur à l'ENSM, Directeur du Centre de CAO et Robotique (CAOR), qui a accepté mon rattachement à son centre en troisième année de thèse. Je le remercie pour sa grande disponibilité et l'intérêt porté à mon travail.

*Fabien MOUTARDE*, Maître-Assistant à l'ENSM, rattaché au CAOR, qui au travers de ses interventions pertinentes lors de mes différentes présentations de passage m'a permis de consolider la rigueur soutenant mon étude.

*Madame Caroline WITTWER*, Directrice-Adjointe du Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations, qui est à l'origine de cette étude. Je la remercie pour sa disponibilité et pour sa clairvoyance. J'en profite pour la gratifier d'avoir accepté l'invitation à être membre de mon jury de thèse

*Monsieur Marc VINCHES*, Enseignant-Chercheur au CMGD et Chef d'équipe de GMH, qui a contribué à l'aboutissement cette thèse. Ses précieux conseils et encouragements ont permis de me concentrer sur l'essentiel.

*Monsieur Denis DARTUS* et *Monsieur Claude TOUZET* les rapporteurs de mon mémoire de thèse. Je vous suis reconnaissant du temps consacré à l'évaluation de mon travail.

*Monsieur Georges-Marie SAULNIER*, Chargé de Recherche au laboratoire Environnements Dynamiques et Territoires de Montagne (EDYTEM), pour avoir accepté de faire partie de mon jury de thèse en tant qu'examinateur.



*Monsieur Pierre-Alain AYRAL*, Enseignant-Chercheur au Laboratoire Génie de l'Environnement Industriel et des risques industriels et naturels (LGEI) et membre de l'équipe Risques Naturels, qui m'a fourni une aide précieuse dans l'élaboration de ma base d'événements de type crues éclair.

*Monsieur Bernard VAYSSADE*, Enseignant-Chercheur au CMGD, qui a toujours su me faire des remarques constructives me permettant d'améliorer mes présentations. Je souhaite qu'il trouve ici toute ma reconnaissance.

*Monsieur Olivier LOUIS* et *Monsieur Réginald SARRALDE*, qui se sont succédés au poste d'animateur du projet BVNE au SCHAPI, ont contribué, par leur promptitude, au bon déroulement de ma thèse.

*L'Association des Thésards de l'EMA (ATHEMA)*, m'a permis de travailler dans un cadre convivial, soucieux du bien-être de chacun de ses membres. Je n'oublie pas mes collègues élèves-chercheurs du centre. Je souhaite qu'ils trouvent ici toute ma reconnaissance.

*Madame Annie MOULIN* de l'Association des Internationaux des Mines d'Alès (AIMA), qui m'a accueilli dans cette association et s'est toujours soucieuse du bon déroulement de mes travaux au laboratoire et plus globalement dans la ville d'Alès.

*La Famille BRUNEL* qui a accepté de me parrainer durant ces trois années de thèse. J'aimerais ici leur présenter toute ma gratitude pour leur soutien indéfectible tout au long de mes travaux.

*L'Association Basket-ball Club – Jeunesse Sportive d'Alès (ABC-JSA)* que j'ai pu intégrer pendant les deux premières années de ma thèse. Je tiens à les remercier pour leurs encouragements.

Enfin, j'aimerais associer à ce moment particulier de ma carrière, ma famille, et tout particulièrement ma mère, pour sa sagesse.

# Plan du mémoire

<b>Plan du mémoire .....</b>	<b>1</b>
<b>Liste des Figures.....</b>	<b>4</b>
<b>Liste des Tableaux .....</b>	<b>6</b>
<b>Introduction .....</b>	<b>7</b>
<b>I Réseaux de neurones pour la prévision .....</b>	<b>9</b>
<b>I.1 Introduction aux réseaux de neurones formels .....</b>	<b>9</b>
I.1.1 Le neurone formel.....	9
I.1.2 Les différents réseaux de neurones .....	10
<b>I.2 Propriétés fondamentales des réseaux de neurones .....</b>	<b>12</b>
I.2.1 L'approximation universelle .....	12
I.2.2 La parcimonie .....	13
I.2.3 De l'approximation de fonction à la modélisation statistique .....	14
<b>I.3 Apprentissage des réseaux de neurones .....</b>	<b>14</b>
I.3.1 Position du problème.....	14
I.3.2 Minimisation de la fonction de coût .....	14
I.3.3 Apprentissage des réseaux dynamiques .....	15
I.3.4 Problème du surajustement .....	16
I.3.5 La sélection de variables.....	17
I.3.6 Les méthodes de régularisation.....	18
I.3.7 Estimation de l'erreur de généralisation : la validation croisée.....	19
<b>I.4 Modélisation de processus dynamiques pour la prévision .....</b>	<b>20</b>
I.4.1 Les représentations d'un système dynamique .....	20
I.4.2 La modélisation pour la prévision .....	25
<b>I.5 Conclusion .....</b>	<b>27</b>
<b>I.6 Références bibliographiques du Chapitre I.....</b>	<b>28</b>
<b>II Prévision des crues éclair du Gardon d'Anduze .....</b>	<b>31</b>
<b>II.1 Introduction .....</b>	<b>31</b>
<b>II.2 Les Gardonnades .....</b>	<b>31</b>
<b>II.3 Etudes hydrologiques sur le bassin versant d'Anduze.....</b>	<b>33</b>
II.3.1 La modélisation hydrologique .....	33
II.3.2 Application de modèles hydrologiques au Gardon d'Anduze.....	35
II.3.3 Synthèse .....	38
<b>II.4 Critères de qualité.....</b>	<b>38</b>
II.4.1 Coefficient de détermination, ou critère de Nash .....	38

II.4.2	Critère de persistance .....	39
II.4.3	Critère de hauteur.....	40
<b>II.5</b>	<b>Projet BVNE Gardon .....</b>	<b>41</b>
II.5.1	Échantillonnage .....	41
II.5.2	Cahier des charges du projet BVNE.....	42
II.5.3	Base de données.....	45
<b>II.6</b>	<b>Conclusion .....</b>	<b>50</b>
<b>II.7</b>	<b>Références bibliographiques du Chapitre II .....</b>	<b>51</b>
<b>III</b>	<b><i>Application des réseaux de neurones à la prévision des crues éclair du bassin versant d'Anduze .....</i></b>	<b>53</b>
<b>III.1</b>	<b>Introduction .....</b>	<b>53</b>
<b>III.2</b>	<b>Mise en œuvre des réseaux de neurones pour la prévision.....</b>	<b>53</b>
III.2.1	Prévision des crues par réseaux de neurones.....	53
III.2.2	Modèle de prédicteur proposé.....	55
<b>III.3</b>	<b>Sélection des variables .....</b>	<b>57</b>
III.3.1	Dimensionnement de la fenêtre temporelle de pluie $f_T$ .....	57
III.3.2	Variables supplémentaires .....	57
<b>III.4</b>	<b>Sélection de modèle par validation croisée.....</b>	<b>59</b>
III.4.1	Application de la validation croisée .....	59
<b>III.5</b>	<b>Impact des méthodes de régularisation actives .....</b>	<b>66</b>
III.5.1	Régularisation par arrêt précoce .....	66
III.5.2	Régularisation par modération de poids .....	84
III.5.3	Régularisation par cumul de l'arrêt précoce et de la modération de poids .....	89
III.5.4	Synthèse de l'application des méthodes de régularisation .....	93
<b>III.6</b>	<b>Conclusion.....</b>	<b>102</b>
<b>III.7</b>	<b>Références bibliographiques du chapitre III .....</b>	<b>103</b>
	<b><i>Conclusion Générale .....</i></b>	<b>105</b>
	<b><i>Annexe A : Méthode de Levenberg-Marquardt.....</i></b>	<b>107</b>
	<b><i>Annexe B : Courbe de tarage .....</i></b>	<b>109</b>
	<b><i>Annexe C : Origine des événements utilisés .....</i></b>	<b>111</b>
	<b><i>Annexe D : Présentation des événements de la base de données.....</i></b>	<b>113</b>
	<b><i>Annexe E : La crue exceptionnelle du 9 septembre 2002.....</i></b>	<b>121</b>
	<b><i>Annexe F : Prévisions « naïves » de la crue de septembre 2002.....</i></b>	<b>125</b>

<b><i>Annexe G : Meilleures prévisions obtenues .....</i></b>	<b><i>127</i></b>
<b><i>Annexe H : Article accepté par la revue « Applied Intelligence ».....</i></b>	<b><i>157</i></b>

# Liste des Figures

Figure 1 : Représentation du neurone formel .....	9
Figure 2 : Un réseau de neurones non bouclé à une couche de neurones cachés (perceptron multicouche) .....	11
Figure 3 : Réseau de neurones bouclé .....	12
Figure 4 : L'approximateur universel : un réseau de neurones non bouclé à $n+1$ entrées, $N_c$ neurones cachés et un neurone de sortie linéaire .....	13
Figure 5 : Apprentissage d'un réseau dynamique avec la méthode des copies .....	16
Figure 6 : Procédure de validation croisée .....	20
Figure 7 : Représentation d'état .....	21
Figure 8: Représentation entrée-sortie .....	22
Figure 9 : Modèle postulé (9-a) et prédicteur neuronal (9-b) associés à l'hypothèse bruit d'état .....	22
Figure 10 : Modèle postulé (10-a) et prédicteur neuronal (10-b) associé à l'hypothèse bruit de sortie .....	24
Figure 11 : Représentation des erreurs de mesure de précipitation vues comme une perturbation d'état .....	25
Figure 12 : Le bassin versant d'Anduze et sa situation en France .....	32
Figure 13 : Carte vigilance crue (exemple pour le 8 octobre 2009) .....	43
Figure 14 : Carte locale des cours d'eau gérés par le SPC Grand Delta (le 8 octobre 2009) .....	44
Figure 15 : Modèle postulé (1-a) de la relation pluie-débit, et prédicteur neuronal dirigé (1-b) associé à l'hypothèse bruit d'état .....	56
Figure 16 : Le type de prédicteur choisi .....	56
Figure 17 : Evolution du cumul des précipitations de l'événement de septembre 2002 à Anduze .....	58
Figure 18 : Humidité ISBA lors de l'événement de septembre 2002 .....	59
Figure 19 : Evolution du score minimum de l'erreur quadratique moyenne pour chaque horizon de prévision. ..	62
Figure 20 : Limnigrammes de test avec la validation croisée seule .....	65
Figure 21 : Prévision sur l'ensemble d'arrêt constitué de 3 événements ( $h_p=2h$ ) .....	68
Figure 22 : Généralisation à l'horizon 2 h pour l'événement de septembre 2002 .....	69
Figure 23 : Meilleurs $R^2$ calculés en fonction de la complexité .....	71
Figure 24 : Limnigrammes de l'événement de 2002 utilisé en arrêt et en test .....	72
Figure 25 : Précipitations et limnigramme de l'événement de septembre 2000 .....	74
Figure 26 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée avec arrêt précoce sur l'ensemble de validation, pour chaque horizon de prévision .....	76
Figure 27 : Limnigrammes de test avec la validation croisée et l'arrêt précoce sur l'ensemble de validation .....	78
Figure 28 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec arrêt précoce sur l'événement 13, pour chaque horizon de prévision .....	81
Figure 29 : Limnigrammes de test avec la validation croisée partielle et l'arrêt précoce sur l'événement 13 .....	83
Figure 30 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec modération des poids .....	86
Figure 31 : Limnigrammes de test obtenus avec la validation croisée partielle et la régularisation par modération des poids .....	88
Figure 32 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec double régularisation .....	91
Figure 33 : Limnigrammes de test avec la validation croisée partielle et la double régularisation .....	93

<i>Figure 34 : Limnigrammes de test sur l'événement 4 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13.....</i>	<i>96</i>
<i>Figure 35 : Limnigrammes de test sur l'événement 4 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13.....</i>	<i>98</i>
<i>Figure 36 : Limnigrammes de test sur l'événement 23 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13.....</i>	<i>100</i>

# Liste des Tableaux

<i>Tableau 1 : Liste des événements et leurs caractéristiques hydrologiques .....</i>	<i>46</i>
<i>Tableau 2 : Classification des événements .....</i>	<i>48</i>
<i>Tableau 3 : Répartition des événements dans les classes.....</i>	<i>48</i>
<i>Tableau 4 : Influence de l'échantillonnage sur l'amplitude et le déphasage du pic principal de l'événement de septembre 2002. ....</i>	<i>50</i>
<i>Tableau 5 : Meilleurs modèles issus de la validation croisée seule .....</i>	<i>63</i>
<i>Tableau 6: Critères de qualité de la prévision avec la validation croisée seule.....</i>	<i>66</i>
<i>Tableau 7 : Critères de qualité de référence .....</i>	<i>73</i>
<i>Tableau 8 : Meilleurs modèles issus de la validation croisée avec arrêt précoce sur l'ensemble de validation ....</i>	<i>75</i>
<i>Tableau 9 : Critères de qualité des prévisions issues de la validation croisée avec arrêt précoce sur l'ensemble de validation, pour chaque horizon de prévision.....</i>	<i>77</i>
<i>Tableau 10 : Meilleurs modèles issus de la validation croisée partielle avec arrêt précoce sur l'événement 13 ..</i>	<i>80</i>
<i>Tableau 11 : Critères de qualité des prévisions issues de la validation croisée partielle avec arrêt précoce sur l'événement 13, pour chaque horizon de prévision.....</i>	<i>82</i>
<i>Tableau 12 : Meilleurs modèles issus de la validation croisée partielle avec la modération des poids .....</i>	<i>85</i>
<i>Tableau 13 : Critères de qualité des prévisions issues de la validation croisée partielle avec modération des poids .....</i>	<i>87</i>
<i>Tableau 14 : Meilleurs modèles issus de la validation croisée partielle avec la double régularisation .....</i>	<i>90</i>
<i>Tableau 15 : Critères de qualité des prévisions issues de la validation croisée avec double régularisation, pour chaque horizon de prévision. ....</i>	<i>92</i>
<i>Tableau 16 : Critères de qualité des prévisions sur l'événement 04.....</i>	<i>97</i>
<i>Tableau 17 : Critères de qualité des prévisions sur l'événement 107.....</i>	<i>99</i>
<i>Tableau 18 : Critères de qualité des prévisions sur l'événement 23.....</i>	<i>101</i>
<i>Tableau 19 : Rapidité de convergence.....</i>	<i>102</i>

# Introduction

Les crues catastrophiques observées ces dernières années (Vaison-la-Romaine 1992, Nîmes 1988, Gardons 2002) ont occasionné des pertes humaines (plus de 100 victimes en 20 ans dans le Sud-Est de la France) et des dégâts considérables (1,2 milliard d'euros de dégâts en 2002), ce qui pose avec une urgence nouvelle le problème de la prévision des crues rapides. Bien que la population concernée par ces événements commence à acquérir une culture du risque, il est nécessaire que les pouvoirs publics mettent en œuvre des moyens pour permettre à la population de se protéger. Face à ce questionnement, le ministère français chargé du développement durable a créé en 2003 le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) qui, par l'intermédiaire des Services de Prévision des Crues, est responsable du service d'alerte mis à jour sur Internet : *vigicrues*. Le Gardon d'Anduze, objet de cette étude, est un petit bassin versant d'environ 500 km<sup>2</sup> situé dans le Sud-Est de la France, qui est sujet à des crues éclair très violentes. Il a été choisi par le SCHAPI comme site pilote pour réaliser une expérience nationale de comparaison de différents modèles numériques. Le BVNE (Bassin Versant Numérique Expérimental) Gardons a donc pour objectif non seulement de développer des modèles de prévision opérationnels, mais également de permettre les comparaisons entre plusieurs modèles fonctionnant à partir de la même base de données. Actuellement, 7 équipes travaillent pour effectuer les prévisions du Gardon d'Anduze ; parmi celles-ci, l'équipe de l'Ecole des Mines d'Alès intervient non seulement en tant que voisine du site pilote, mais également pour appliquer les méthodes de l'apprentissage statistique à la modélisation des processus hydrologiques afin d'estimer les crues, les étiages et les réserves d'eau souterraine.

Dans ce contexte, ce mémoire se propose de présenter les travaux effectués pour réaliser la prévision des crues du Gardon d'Anduze à Anduze, jusqu'à un horizon de prévision de l'ordre de grandeur du temps de concentration, en l'absence de prévision de pluie, par des réseaux de neurones formels.

La première partie de ce mémoire s'attache donc à présenter les réseaux de neurones utilisés pour la prévision du comportement d'un processus dynamique non linéaire. Seront abordées les propriétés fondamentales qui justifient l'utilisation de ces réseaux de neurones (l'approximation universelle et la parcimonie), ainsi que les méthodes connues pour éviter le piège du surajustement. Nous étudierons particulièrement la modélisation entrée-sortie des processus dynamiques, et les conséquences du choix du modèle postulé (modèle avec bruit d'état ou bruit de sortie). Des illustrations dans le domaine de l'hydrologie seront présentées.

La deuxième partie de ce mémoire présente le cours d'eau qui constitue l'objet de l'étude : le Gardon d'Anduze et ses Gardonnades. Ce petit bassin versant est sujet, durant l'automne, à des orages extrêmement violents dont l'intensité peut atteindre presque 24 mm de pluie en une ½ heure, et dont le cumul des précipitations en 24 heures peut dépasser la valeur du cumul de pluie tombée en une année en région parisienne ! Comme on peut l'imaginer, lors de ces événements, le moindre ruisseau se transforme en torrent tumultueux et les rivières sortent de leur lit en emportant ponts, routes, voitures et même l'instrumentation destinée à mesurer les hauteurs d'eau. Il est utile de préciser que les mesures de terrain sont alors extrêmement difficiles à réaliser autant parce que les sites sont coupés du reste du monde que



parce que ces pluies sont extrêmement hétérogènes dans le temps et dans l'espace. Néanmoins pour assurer le service *vigicrues*, le Service de Prévision des Crues « Grand Delta », chargé de la zone des Gardons, dispose des mesures de 6 pluviomètres répartis sur le bassin versant, qui télétransmettent le cumul des précipitations toutes les 5 minutes, ainsi que des stations de mesure de la hauteur d'eau qui télétransmettent également les mesures toutes les 5 minutes.

Une quinzaine d'événements ont donc ainsi été enregistrés depuis 1995 et vont être mis à profit pour concevoir et réaliser le prédicteur des hauteurs d'eau.

Le premier chapitre de ce mémoire ayant présenté la nécessité d'un contrôle de la complexité du modèle neuronal pour garantir une généralisation satisfaisante, le troisième chapitre détaille la méthodologie utilisée pour permettre l'apprentissage sur une base de données très réduite tout en conservant des événements inutilisés de façon à assurer un test de la qualité du modèle qui soit totalement indépendant des procédures d'apprentissage et de sélection de modèle. Les méthodes de régularisation présentées dans le premier chapitre sont mises en œuvre et adaptées de manière originale, de façon à privilégier le phénomène de crue intense. Enfin, ce mémoire présente les limnigrammes prévus pour l'événement majeur de septembre 2002, et démontre ainsi que le niveau d'eau à Anduze peut être prévu avec une assez bonne fiabilité jusqu'à l'horizon de 3 h, permettant ainsi d'alerter les populations de manière précoce.

# I Réseaux de neurones pour la prévision

## I.1 Introduction aux réseaux de neurones formels

Les réseaux de neurones formels constituent une famille de modèles non linéaires paramétrés qui sont mis à profit, dans ce travail, pour des tâches de régression. Si les termes *neurones* ou *apprentissage* sont le reflet de l'origine biologique de ces fonctions, l'approche choisie s'inscrit néanmoins clairement dans le cadre de l'apprentissage artificiel. Dans toute la suite, l'expression « réseaux de neurones » désignera les réseaux de neurones formels.

La prévision par réseaux de neurones s'inscrit dans le cadre de la modélisation de processus de type « boîte noire ». Dans cette étude, le processus étudié est celui de la génération d'une crue éclair du fait de pluies intenses.

Après une présentation des modèles de neurones et de réseaux utilisés en identification de processus, nous nous intéresserons aux propriétés qui fondent l'utilisation de ces réseaux : l'approximation universelle et la parcimonie. Nous considérons l'apprentissage comme le calcul d'une régression non linéaire, cette analogie avec les statistiques nous conduisant à introduire la problématique de la régularisation afin d'optimiser les performances de généralisation. La conception d'un réseau de neurones pour la prévision est ensuite présentée et nous nous intéresserons en particulier aux différents modèles de prévision découlant d'hypothèses différentes sur la manière qu'ont les perturbations d'agir sur le processus. Enfin, une illustration de ces considérations sera décrite.

### I.1.1 Le neurone formel

Un neurone formel réalise une fonction algébrique non linéaire, paramétrée et à valeurs bornées, de ses variables d'entrée :

$$y = f(x_0, \dots, x_n; w_0, \dots, w_n)$$

où les  $\{x_j, 0 \leq j \leq n\}$  sont les variables et les  $\{w_j, 0 \leq j \leq n\}$  sont des paramètres.

Un neurone formel est représenté graphiquement sur la Figure 1.

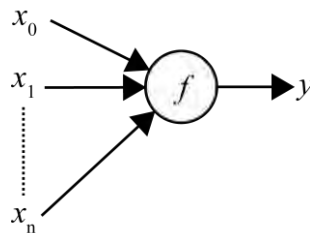


Figure 1 : Représentation du neurone formel

Lorsque la fonction  $f(\cdot)$  est l'identité, on dit que le neurone est linéaire.

Le paramétrage de la fonction  $f(\cdot)$  est souvent réalisé d'une des deux manières suivantes :

- Les paramètres sont liés à la non-linéarité du neurone : ils sont associés à la fonction  $f(\cdot)$  elle-même. Ce type de paramétrage permet de définir les réseaux de fonctions radiales [POWELL 1985] ou d'ondelettes [MEYER 1990].
- Les paramètres sont liés aux variables du neurone : celui-ci effectue une combinaison de ses variables  $\{x_j\}$  pondérées par les paramètres  $\{w_j\}$ , puis transforme non

linéairement cette combinaison. Les paramètres  $\{w_j\}$  sont parfois appelés *poids synaptiques*. La combinaison, souvent linéaire, est désignée par le terme de *potentiel*. Le potentiel, noté  $v$ , est alors la somme pondérée des variables du neurone à laquelle s'ajoute un terme constant appelé *biais*, effectuant un décalage affine :

$$v = w_0 + \sum_{j=1}^n w_j x_j \text{ ou encore } v = \sum_{j=0}^n w_j x_j \text{ où } x_0=1.$$

La fonction  $f(\cdot)$ , appliquée au potentiel  $v$  en vue d'évaluer la sortie  $y$  du neurone, est appelée la *fonction d'activation* :

$$y = f(v) = f\left(\sum_{j=0}^n w_j x_j\right).$$

Tout au long de ce travail, les neurones sont définis par la relation précédente.

Il existe plusieurs fonctions d'activation. Les plus fréquemment utilisées sont :

- La fonction d'activation de Heaviside ou l'échelon : cette fonction limite la sortie du neurone formel à 0 si le potentiel  $v$  du neurone est négatif, ou à 1 si  $v$  est positif ou nul. L'utilisation de l'échelon permet l'implémentation de neurones binaires servant à la classification.
- La fonction d'activation linéaire : elle est utilisée pour le neurone de sortie d'un réseau destiné à modéliser une grandeur qui n'est pas bornée par 0 ou 1.
- La fonction d'activation *sigmoïde* : elle possède plusieurs variantes ; la plus utilisée est la tangente hyperbolique. L'intérêt de cette fonction est qu'elle est continue, dérivable, et qu'elle n'est pas locale puisqu'elle ne tend pas vers zéro lorsque la valeur absolue du potentiel tend vers l'infini.

### 1.1.2 Les différents réseaux de neurones

Un réseau de neurones est une composition des fonctions neurones définies dans la section précédente. On distingue généralement deux types de neurones :

- Les neurones *cachés* : ils sont appelés ainsi parce que leurs sorties ne sont pas des sorties du réseau.
- Les neurones de sortie : leurs sorties sont les sorties du réseau.

Dans ce cadre, il existe deux types d'architectures de réseaux de neurones :

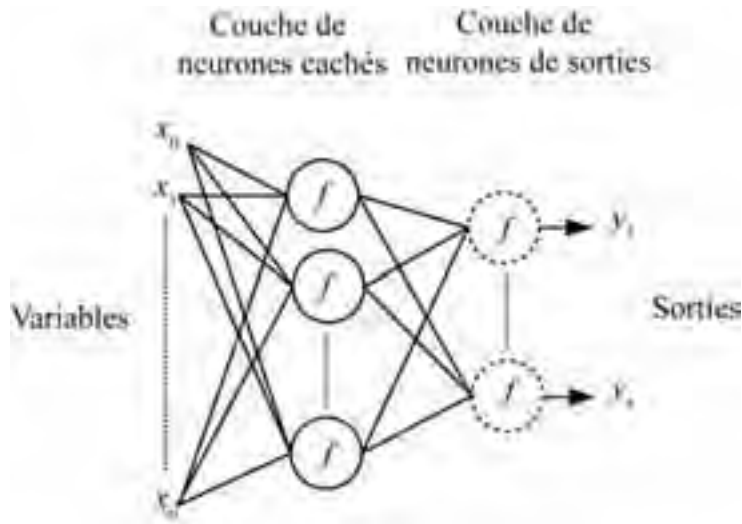
- Les réseaux de neurones statiques.
- Les réseaux de neurones dynamiques.

#### 1.1.2.a Les réseaux de neurones statiques

Un réseau de neurones à  $n$  variables est dit statique s'il réalise une (ou plusieurs) fonctions de ses  $n$  variables et d'elles seules. En conséquence, si ces variables ne varient pas dans le temps, la ou les sorties du réseau ne varient pas. En d'autres termes, le temps ne joue aucun rôle fonctionnel dans un tel réseau.

Un réseau statique à une couche de neurones cachés (perceptron multicouche) est représenté sur la Figure 2. Le graphe dont les nœuds sont les neurones, et les arêtes orientées les

« connexions » entre ceux-ci, est acyclique. C'est pourquoi un tel réseau est appelé *réseau non bouclé*.



**Figure 2 : Un réseau de neurones non bouclé à une couche de neurones cachés (perceptron multicouche)**

Dans le cas où les variables d'entrée sont des signaux échantillonnés, les réseaux statiques représentent des filtres numériques transverses non linéaires et non récurrents :

$$y(k) = \varphi_{RN} \mathbf{x}(k), \dots, \mathbf{x}(k-n)$$

où :

- $y(k)$  est la sortie du filtre à l'instant discret  $k$ .
- $\mathbf{x}(k)$  est le vecteur des variables à l'instant  $k$ .
- $\varphi_{RN}$  est la transformation non linéaire effectuée par le filtre transverse.

#### 1.1.2.b Les réseaux de neurones dynamiques

Un réseau de neurones à  $n$  variables est dit dynamique s'il réalise une (ou plusieurs) fonction de ses  $n$  variables (dites *variables exogènes*) et, en outre, d'une ou plusieurs valeurs passées des sorties de neurones du réseau (*variables récurrentes*). En conséquence, même si les  $n$  variables ne varient pas dans le temps, la ou les sorties du réseau sont susceptibles de varier. Un tel réseau est donc bien un système dynamique.

La Figure 3 présente un exemple de réseau de neurones dynamique. Ce réseau possède autant de cycles que de neurones cachés. Ces cycles partent tous d'un neurone caché donné, en passant par l'unique neurone de sortie, puis reviennent au neurone caché de départ avec un retard d'une période d'échantillonnage.

Le graphe de ses connexions, défini dans la section précédente, est cyclique ; c'est pourquoi ce réseau est appelé *réseau bouclé*. Pour qu'il soit causal, il faut que chaque boucle contienne au moins un retard unité. Il réalise alors une ou plusieurs équations récurrentes.

Un réseau bouclé à une sortie, dont les variables d'entrée sont constituées de variables exogènes et de valeurs passées de la sortie, est également appelé « filtre récursif non linéaire ».

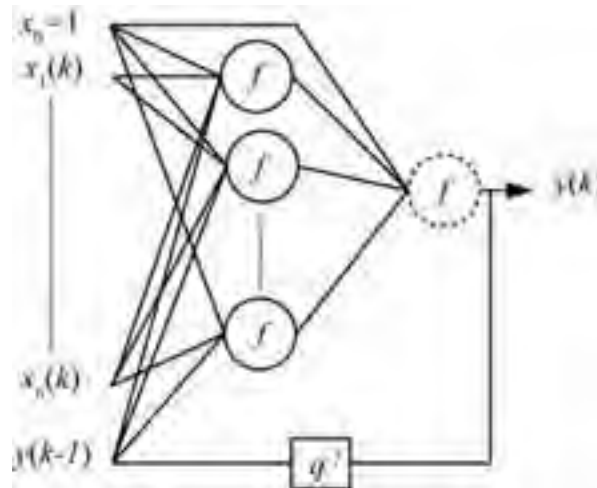


Figure 3 : Réseau de neurones bouclé.

Les travaux de [NERRAND *et al.* 1993] ont établi que tout réseau de neurones bouclé, aussi complexe soit-il, peut être mis sous une forme particulière, appelée forme canonique, comportant un réseau de neurones non bouclé dont certaines sorties sont ramenées aux entrées par des bouclages de retard unité. Cette forme canonique est par conséquent constituée d'un graphe acyclique, et de connexions à retard unité reliant certaines sorties de ce graphe aux entrées du réseau.

## 1.2 Propriétés fondamentales des réseaux de neurones

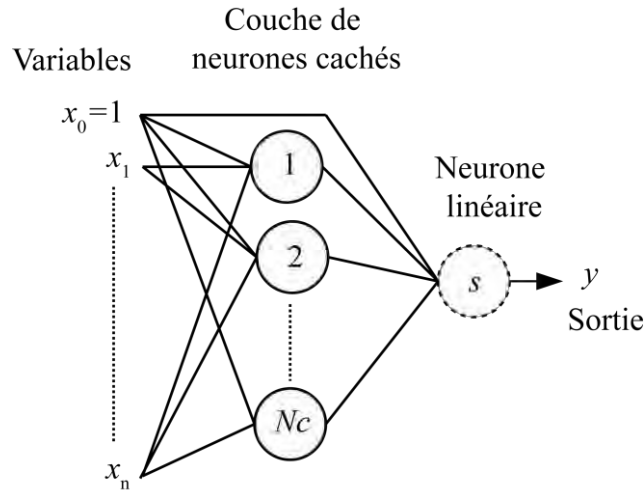
Les réseaux de neurones constituent une famille de fonctions non linéaires paramétrées, que l'on met en œuvre pour des tâches de modélisation, de prédiction, et de classification.

### 1.2.1 L'approximation universelle

[CYBENKO 1989] et [FUNAHASHI 1989] ont démontré la propriété d'approximation universelle des réseaux de neurones que nous énonçons ci-après :

« Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation bornée, et un neurone de sortie linéaire. » [HORNIK *et al.* 1989], [HORNIK *et al.* 1990], [HORNIK 1991]

Cette propriété fondamentale est un théorème d'existence. Elle ne simplifie pas pour autant la tâche consistant à identifier les paramètres du réseau. Toutefois, elle permet de fixer l'architecture de réseaux pouvant réaliser l'identification de toute fonction non linéaire.



**Figure 4 : L'approximateur universel : un réseau de neurones non bouclé à  $n+1$  entrées,  $N_c$  neurones cachés et un neurone de sortie linéaire**

La Figure 4 présente l'approximateur universel : la sortie  $y$  de ce réseau s'écrit de la façon suivante :

$$y = w_{s,0} + \sum_{j_2=1}^{N_c} \left[ w_{s,j_2} f \left( w_{j_2,0} + \sum_{j_1=1}^n w_{j_2,j_1} x_{j_1} \right) \right]$$

où :

- $f$  est la fonction d'activation sigmoïde des neurones de la couche cachée,
- $w_{s,0}$  est le paramètre reliant l'entrée constante d'indice 0 au neurone de sortie d'indice  $s$ ,
- $w_{j_2,0}$  est le paramètre reliant l'entrée constante d'indice 0 au neurone de la couche cachée d'indice  $j_2$ , avec  $j_2 \in [1, N_c]$ ,
- $w_{s,j_2}$  est le paramètre reliant le neurone de la couche cachée d'indice  $j_2$  au neurone de sortie d'indice  $s$ ,
- $w_{j_2,j_1}$  est le paramètre reliant la variable d'entrée  $x_{j_1}$  ( $j_1 \in [1, n]$ ) au neurone de la couche cachée d'indice  $j_2$ .

Il est important de noter que la fonction réalisée par un tel réseau est une fonction non linéaire de ses variables et de ses paramètres.

### 1.2.2 La parcimonie

Par ailleurs, [BARRON 1993] montre que : si l'approximation dépend des paramètres ajustables de manière non linéaire, elle est plus parcimonieuse que si elle dépend linéairement des paramètres. Plus précisément, on montre que le nombre de paramètres, pour une précision donnée, croît exponentiellement avec le nombre de variables dans le cas des approximateurs linéaires par rapport à leurs paramètres, alors qu'il croît linéairement avec ce nombre pour les approximateurs non linéaires par rapport à leurs paramètres. La parcimonie est donc d'autant plus avantageuse que le nombre de variables du modèle est grand.

Les réseaux de neurones qui ont été décrits dans la section précédente étant non linéaires par rapport à leurs paramètres, ils sont plus parcimonieux que les approximateurs universels linéaires par rapport à leurs paramètres, tels que les polynômes par exemple.

### 1.2.3 De l'approximation de fonction à la modélisation statistique

Les réseaux de neurones ne sont généralement pas utilisés pour réaliser une approximation uniforme de fonctions, mais pour réaliser une tâche de régression non linéaire.

Dans ce cadre, on modélise les  $m$  observations disponibles de la grandeur à modéliser  $y^p(\mathbf{x})$  comme des réalisations d'une variable aléatoire  $Y$ , fonction de  $\mathbf{x}$ , qui est la somme d'une fonction certaine  $h(\mathbf{x})$  et d'une variable aléatoire  $B$  d'espérance mathématique nulle

$$Y(\mathbf{x}) = h(\mathbf{x}) + B.$$

Cette dernière variable modélise l'ensemble des perturbations susceptibles d'affecter les mesures. La fonction  $h(\mathbf{x})$ , appelée *fonction de régression*, est donc égale à l'espérance mathématique de  $Y$  conditionnée à  $\mathbf{x}$  :

$$h(\mathbf{x}) = E(Y | \mathbf{x}).$$

On cherche alors le réseau de neurones qui constitue la meilleure approximation possible de la fonction inconnue  $h(\mathbf{x})$  compte tenu des données disponibles, dans le domaine de variation des variables  $\mathbf{x}$ .

## 1.3 Apprentissage des réseaux de neurones

L'apprentissage des réseaux de neurones consiste à estimer les paramètres  $\mathbf{w}$  du modèle, de telle manière qu'une distance convenablement choisie entre les observations appartenant à un ensemble de données, l'*ensemble d'apprentissage*, et les prédictions du modèle pour ces mêmes données soit aussi petite que possible.

### 1.3.1 Position du problème

On dispose d'un ensemble de  $N$  mesures  $\{y^k\}$  ( $k = 1, \dots, N$ ) de la grandeur à modéliser, et des valeurs correspondantes des  $n$  variables  $\{\mathbf{x}^k\} = \{[x_1^k, \dots, x_n^k]\}$ . Cet ensemble de  $N$  couples d'entrées-sorties constitue l'ensemble d'apprentissage.

On cherche les valeurs des paramètres pour lesquelles une fonction de coût, représentative des différences entre les valeurs mesurées et les valeurs calculées par le modèle, est minimale. La fonction la plus utilisée est la fonction de coût des moindres carrés :

$$J(\mathbf{w}) = \sum_{k=1}^N (y^k - y(\mathbf{x}^k, \mathbf{w}))^2$$

où  $y(\mathbf{x}^k, \mathbf{w})$  est la valeur de la sortie du réseau pour les variables  $\mathbf{x}^k$ .

### 1.3.2 Minimisation de la fonction de coût

Le modèle étant non linéaire en ses paramètres, la fonction de coût n'est pas quadratique en les paramètres. La méthode des moindres carrés n'est donc pas applicable. En conséquence, on a recours à des méthodes itératives d'optimisation de la fonction de coût.

Toutes les méthodes d'optimisation utilisent le gradient de la fonction de coût. La première étape de l'apprentissage d'un réseau de neurones consiste donc à calculer le gradient de la fonction de coût, à l'aide de l'algorithme de *rétropropagation* [RUMELHART *et al.* 1986].

Une fois le gradient calculé, on met en œuvre un algorithme itératif de modification des paramètres. Parmi ces derniers, on distingue les méthodes itératives du premier ordre et les méthodes du second ordre.

Les méthodes du premier ordre modifient itérativement les paramètres de manière proportionnelle au gradient de la fonction de coût, avec un coefficient de proportionnalité fixe ou variable au cours du déroulement de l'optimisation. En d'autres termes, l'extrémité du vecteur des paramètres se déplace, à chaque itération, dans la direction du gradient de la fonction de coût.

Dans les méthodes du second ordre, la direction de déplacement du vecteur des paramètres est obtenue par une transformation linéaire du gradient de la fonction de coût, transformation qui fait intervenir la matrice des dérivées secondes de la fonction de coût par rapport aux paramètres (matrice hessienne). Ces méthodes sont beaucoup plus efficaces que les méthodes du premier ordre. Le choix entre les diverses méthodes du second ordre dépend notamment du nombre de paramètres des modèles étudiés. Compte tenu de la taille modeste de nos modèles, nous avons choisi de mettre en œuvre la méthode de Levenberg-Marquardt qui utilise une approximation de la matrice hessienne calculée à l'aide des produits des dérivées premières [BENDER 1996] [PRESS 1992] (Annexe A.). Cette méthode est avantageuse car il n'est pas nécessaire de calculer les dérivées secondes de la fonction de coût.

### **1.3.3 Apprentissage des réseaux dynamiques**

Lorsque le modèle de réseau est dynamique (récurrent), il est avantageux d'effectuer l'apprentissage sur la représentation canonique du modèle bouclé [NERRAND *et al.* 1993]. Comme la représentation canonique présente le modèle sous la forme d'une partie acyclique constituée d'un réseau non bouclé et d'une partie temporelle comprenant les bouclages et les retards déportés à l'extérieur du réseau non bouclé, la prise en compte du temps dans le réseau se fait en « coupant » ces bouclages, et en dupliquant le réseau non bouclé en autant de *copies* dans le temps que l'on veut remonter dans le passé. L'apprentissage peut alors être réalisé comme pour un réseau non bouclé au travers des différentes copies comme représenté par la flèche sur la Figure 5, avec autant de couches cachées qu'il y a de copies. Par exemple, la Figure 5 représente un réseau possédant un bouclage, ainsi déplié pour une profondeur de 2 périodes. Il est important de noter que, comme à l'issue de l'apprentissage il n'existe qu'un seul réseau, les paramètres des copies doivent être égaux entre eux à la fin de l'apprentissage, la technique des *poids partagés* est utilisée afin de satisfaire à cette contrainte.



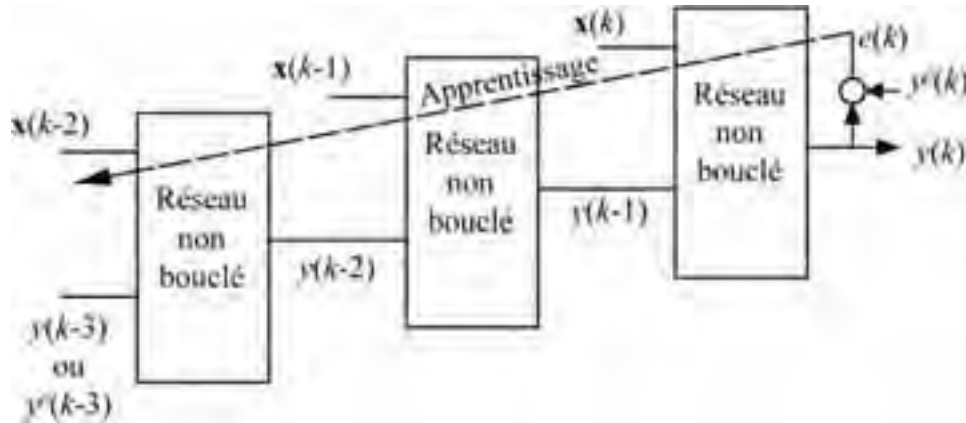


Figure 5 : Apprentissage d'un réseau dynamique avec la méthode des copies

Les variables exogènes, représentées par le vecteur  $\mathbf{x}(k-c)$  sont appliquées comme les variables de chaque copie  $c$ , et la sortie  $y(k-c)$  calculée par la copie  $c$  est appliquée à la copie suivante ; la dynamique du processus est ainsi prise en considération. Se pose le problème de l'initialisation de la copie correspondant au temps le plus ancien : si la variable correspondant à la sortie du processus au temps précédent est choisie ( $y^p(k-3)$  sur la Figure 5), l'apprentissage est dit *semi-dirigé* par le processus.

### I.3.4 Problème du surajustement

Comme nous l'avons indiqué dans la section I.2.3, l'objectif d'un modèle n'est pas de réaliser l'approximation d'une fonction connue, mais de chercher la meilleure approximation possible de la fonction de régression, compte tenu des données disponibles. Cette approximation est obtenue en postulant une fonction paramétrée (un réseau de neurones par exemple), et en estimant les valeurs de ses paramètres pour qu'elle s'ajuste aux données de l'ensemble d'apprentissage. Il faut donc être attentif au fait qu'un modèle trop complexe risque de s'ajuster, non pas à la fonction de régression, mais à la somme de celle-ci et du bruit aléatoire présent dans les données. Un modèle qui est ajusté au bruit est évidemment incapable de généraliser correctement, puisqu'il dépend de la réalisation particulière du bruit présent dans les données d'apprentissage. On dit qu'un tel modèle est *surajusté*.

À l'inverse, un modèle de complexité insuffisante risque de ne pas avoir la « souplesse » suffisante pour s'ajuster à la fonction de régression. Il faut donc trouver la complexité appropriée : celle qui permet à la fonction postulée de constituer une bonne approximation de la fonction de régression sans s'ajuster au bruit présent dans les données.

Ce problème, appelé *dilemme biais-variance*, constitue un problème central de l'apprentissage statistique. Il est traduit mathématiquement de la manière suivante : supposons, comme nous l'avons indiqué dans la section I.2.3, que les observations soient la somme d'une fonction certaine (fonction de régression) et d'une variable aléatoire  $B$  d'espérance mathématique nulle et de variance  $\sigma^2$ , qui modélise toutes les perturbations qui affectent la mesure :

$$Y(\mathbf{x}) = h(\mathbf{x}) + B.$$

Alors l'erreur de prédiction théorique, pour un vecteur  $\mathbf{x}$  donné, c'est-à-dire l'espérance mathématique du carré de la différence entre l'observation et la prédiction du modèle correspondante, calculée sur tous les ensembles d'apprentissage possible, est la somme de trois termes [GEMAN *et al.* 1992] :

- un terme qui ne dépend que des données : la variance  $\sigma^2$  de la variable  $B$ , c'est-à-dire la variance du « bruit » présent dans les données,
- la variance du modèle, calculée sur tous les ensembles d'apprentissage possibles,
- le « biais » du modèle, qui est le carré de l'espérance mathématique de la différence entre la fonction de régression et la prédiction du modèle, calculée sur tous les ensembles d'apprentissage possibles.

Ces trois termes étant positifs, il en résulte que l'erreur d'un modèle obtenu par apprentissage ne peut être inférieure à l'incertitude sur les données (variance  $\sigma^2$  de la variable  $B$ ).

De plus, le biais et la variance varient en sens inverse en fonction de la complexité du modèle : le biais diminue lorsque la complexité augmente (le modèle s'adapte de plus en plus précisément aux données de l'ensemble d'apprentissage), mais la variance augmente (le modèle est de plus en plus sensible aux détails de l'ensemble d'apprentissage). La somme des deux termes présente donc un minimum pour une certaine complexité, qui constitue le meilleur compromis entre le biais et la variance. Au-delà de cette complexité, le modèle est surajusté.

Ces résultats sont asymptotiques, c'est-à-dire qu'ils sont exacts dans la limite des très grands ensembles de données. Ils sont d'une grande importance théorique, mais ils n'ont pas de valeur opérationnelle.

Dans la pratique, la procédure normale de conception de modèle par apprentissage consiste à concevoir des modèles de complexité croissante, à estimer l'erreur de généralisation de chacun d'eux, et à sélectionner le modèle qui possède les meilleures propriétés de généralisation, donc qui réalise le meilleur compromis entre le biais et la variance. Ce problème est d'autant plus difficile que l'ensemble d'apprentissage est restreint [GALLINARI *et al.* 1999].

Deux types de méthodes permettent d'éviter le surajustement :

- Les méthodes passives : elles permettent d'éliminer, parmi plusieurs modèles de complexités différentes, ceux qui sont susceptibles d'être surajustés. Comme méthode passive, nous présenterons la validation croisée [SAARINEN *et al.* 1993], [ZHOU *et al.* 1998].
- Les méthodes actives : elles évitent le surajustement du réseau en limitant l'amplitude des paramètres lors de l'apprentissage du réseau. Les méthodes de régularisation comme l'arrêt précoce ou la pénalisation des poids permettent de trouver une solution particulière [TIKHONOV *et al.* 1977], [POGGIO *et al.* 1985].

### **1.3.5 La sélection de variables**

La sélection de variables consiste à choisir un sous-ensemble de variables parmi les variables candidates choisies par les experts comme pouvant être pertinentes pour la modélisation du processus. Si l'on sait bien quelles sont les variables qui influencent le processus, la sélection de variables peut être faite uniquement de manière experte, à partir des connaissances. Si le processus est moins bien connu, ou si plusieurs variables comportent la même information, mais au travers de méthodes d'observation différentes, il est alors nécessaire de mettre en œuvre un processus de sélection de variables. Ce dernier a pour objet de diminuer la complexité du modèle afin d'améliorer ses propriétés de généralisation.

Un processus de sélection de variables nécessite de mettre en œuvre :

- la définition d'un critère de pertinence, qui servira à ordonner les variables afin d'examiner leur importance ; ceci peut être fait par une orthogonalisation de Gram-Schmidt ou par des méthodes heuristiques [YACOUB *et al.* 2001]
- la définition d'un seuil qui permet d'accepter ou de rejeter la variable considérée.

Notons que l'ensemble des variables sélectionnées n'est en général pas unique et qu'il peut en exister plusieurs de performances équivalentes. En tout état de cause, compte tenu de la complexité combinatoire que ceci entraînerait, il est bien souvent impossible de tester tous les sous-ensembles possibles de variables.

### 1.3.6 Les méthodes de régularisation

#### 1.3.6.a La régularisation par modération des poids

Si un modèle est trop complexe, l'apprentissage conduit à des valeurs des paramètres qui sont très grandes ; réciproquement, la présence de paramètres dont les valeurs sont très grandes est un symptôme de surajustement. Ainsi, [BARTLETT 1997] montre que, si un grand réseau est utilisé, et si l'algorithme d'apprentissage trouve une erreur quadratique moyenne faible avec des paramètres de valeurs absolues faibles, alors les performances en généralisation dépendent de la norme des poids plutôt que de leur nombre.

La régularisation par modération des poids (*weight decay*), empêche les paramètres du réseau de prendre des valeurs exagérément grandes, donc évite la saturation des fonctions d'activation sigmoïdes.

À cette fin, on ajoute à la fonction de coût des moindres carrés  $J$  un terme de régularisation proportionnel à la somme des carrés des paramètres :

$$J' = J + \frac{\alpha}{2} \sum_{j=1}^z w_j^2$$

où :

- $z$  est le nombre de paramètres du réseau,
- $\alpha$  est un hyperparamètre qui détermine l'importance relative des deux termes dans la fonction de coût  $J'$ .

Tous les paramètres des différentes couches ne jouant pas le même rôle lors de l'apprentissage, il peut être efficace d'affecter des hyperparamètres différents aux différents types de paramètres :

$$J' = J + \frac{\alpha_0}{2} \sum_{i \in \mathbf{w}_0} w_i^2 + \frac{\alpha_1}{2} \sum_{i \in \mathbf{w}_1} w_i^2 + \frac{\alpha_2}{2} \sum_{i \in \mathbf{w}_2} w_i^2$$

où

- $\mathbf{w}_0$  est le vecteur des paramètres reliant le biais aux neurones cachés,
- $\mathbf{w}_1$  est le vecteur des paramètres reliant les variables d'entrée aux neurones cachés,
- $\mathbf{w}_2$  est le vecteur des paramètres reliés au neurone de sortie (biais et paramètres reliant les sorties de la couche cachée).

La détermination des trois nouveaux hyperparamètres  $\alpha_0, \alpha_1, \alpha_2$  peut être effectuée en se fondant sur des hypothèses et calculs statistiques [MCKAY 1992], mais cette détermination peut se faire de façon heuristique de manière satisfaisante [DREYFUS *et al.* 2008].

#### 1.3.6.b L'arrêt précoce

La méthode de régularisation par arrêt précoce consiste à arrêter la phase d'apprentissage dès l'instant où la généralisation commence à se dégrader. Le critère d'arrêt consiste donc à disposer d'un ensemble de données, distinct de l'ensemble d'apprentissage, *l'ensemble d'arrêt*, sur lequel on surveille l'évolution de la fonction de coût en vue d'arrêter la phase d'apprentissage lorsque le coût calculé sur ce nouvel ensemble atteint son minimum.

[SJÖBERG *et al.* 1995] ont prouvé théoriquement que l'arrêt précoce est équivalent à l'introduction, dans la fonction de coût, d'un terme de pénalisation de l'amplitude des paramètres. Il justifie donc que l'arrêt précoce puisse être considéré comme une méthode de régularisation.

### 1.3.7 Estimation de l'erreur de généralisation : la validation croisée

Comme nous l'avons vu précédemment, il n'est pas possible de sélectionner les modèles à partir de l'erreur d'apprentissage, car la différence entre l'erreur d'apprentissage et l'erreur de généralisation croît avec la complexité. En d'autres termes, **l'erreur sur l'ensemble d'apprentissage ne permet pas, en général, d'estimer l'erreur de généralisation.**

La validation croisée permet d'estimer l'erreur de généralisation du modèle à partir d'un ensemble de données qui n'est pas utilisé pour l'estimation des paramètres. Ce nouvel ensemble définit l'ensemble de validation. Dans cette procédure, on sépare l'ensemble des données disponibles en deux sous-ensembles : un ensemble d'apprentissage-validation et un ensemble de test. L'ensemble d'apprentissage-validation est lui-même séparé en  $D$  sous-ensembles disjoints contenant le même nombre d'exemples [STONE 1974]. Puis, pour une complexité de modèle donnée, on effectue  $D$  fois l'union de  $D-1$  sous-ensembles, que l'on utilise pour l'apprentissage de modèles, ce qui permet d'affecter à chaque fois un sous-ensemble d'exemples à la validation (Figure 6). On peut ainsi calculer le score de validation croisée  $S$  issu des  $D$  sous-ensembles affectés successivement à la validation ( $E_v$  pour dire évaluation moyenne) :

$$S = \sqrt{\frac{1}{D} \sum_{E_p=1}^D J(E_p)^2},$$

$$\text{avec } J(E_p) = \frac{1}{N_p} \sum_{m=1}^{N_p} (y^m - \mathbf{x}^m \cdot \mathbf{w})^2,$$

où :

- $E_p$  est le sous-ensemble de données numéro  $p$  ( $p = 1$  à  $D$ ) comportant  $N_p$  exemples,
- $J(E_p)$  est la fonction de coût calculée sur le sous-ensemble  $E_p$ , en l'occurrence l'erreur quadratique moyenne minimale sur l'ensemble de validation en faisant varier l'initialisation des paramètres du réseau,
- $y^m$  et  $\mathbf{x}^m \cdot \mathbf{w}$  sont les grandeurs respectivement mesurée et estimée.

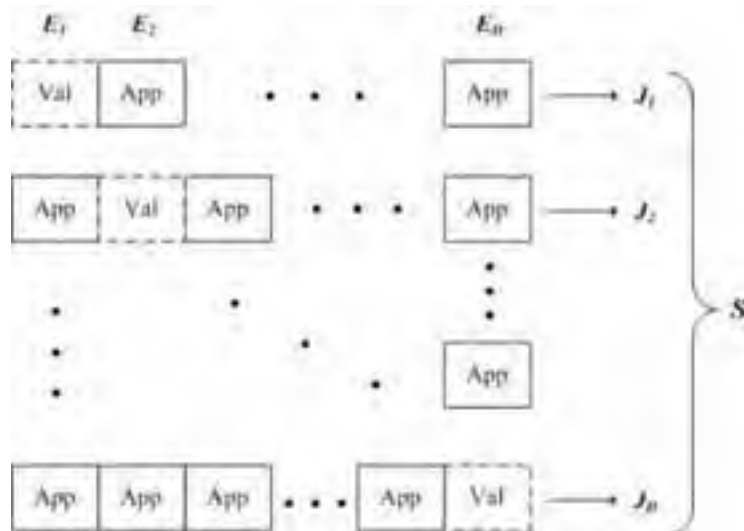


Figure 6 : Procédure de validation croisée

La procédure de validation croisée est appliquée à des modèles de complexité croissante, jusqu'à la détection du surajustement, qui se traduit par le fait que le score de validation croisée augmente, alors que l'erreur sur l'ensemble d'apprentissage diminue. On sélectionne alors le réseau pour lequel la plus petite valeur du score de validation croisée a été obtenue [DREYFUS *et al.* 2008].

La sélection de modèle par validation croisée fixant  $D$  égal à  $N - 1$  où  $N$  est le nombre d'exemples disponibles – est appelée *Leave-One-Out*. C'est une technique très lourde en temps de calcul, mais [VAPNIK 1995] a démontré que le score de validation croisée est un estimateur non biaisé de l'erreur de généralisation.

Une fois la complexité optimale choisie, l'apprentissage d'un modèle possédant cette complexité est effectué à partir de toutes les données de l'ensemble d'apprentissage-validation. Les performances de ce modèle sont estimées sur l'ensemble de test, dont les éléments n'ont jamais été utilisés auparavant, ni pour l'apprentissage ni pour la sélection du modèle.

## 1.4 Modélisation de processus dynamiques pour la prévision

La modélisation d'un processus dynamique nécessite de réaliser des hypothèses sur le type de fonction qui régit son comportement, afin de décrire ce dernier par un modèle mathématique. On peut ainsi proposer un ou plusieurs *modèles hypothèses* ou *modèles postulés* qui seront construits lors de l'identification, puis évalués lors de la phase de validation.

### 1.4.1 Les représentations d'un système dynamique

#### 1.4.1.a Modèles boîtes noires et modèles de connaissance

La recherche préliminaire de connaissance sur le système étudié conduit à déterminer le type de modélisation choisie. Lorsque l'on dispose de suffisamment de connaissances sur le processus, en particulier sur ses fonctions internes, on élaborera plutôt un modèle de connaissance. À l'opposé, si l'on ne dispose que des signaux d'entrée et de sortie et que les

fonctions internes sont inconnues, on conçoit un modèle *boîte noire*, qui peut être décrit par un modèle de type *entrée-sortie* ou de type *état*.

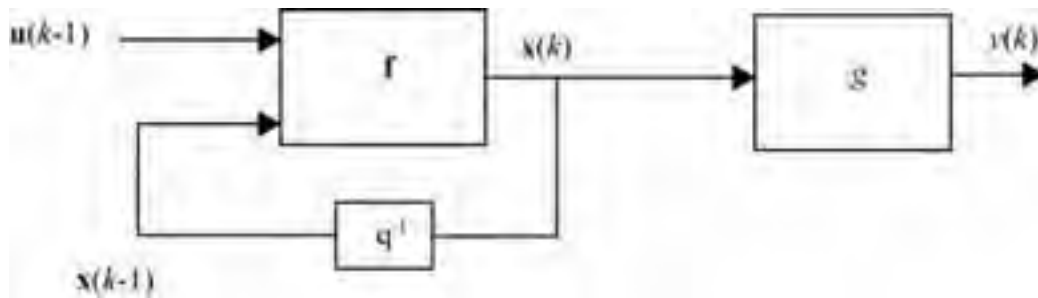
Un système échantillonné est à représentation d'état (Figure 7) si l'on peut déterminer un vecteur  $\mathbf{x}(k)$ , dit de *variables d'état*, de dimension minimale, de sorte que son expression, dans le cas d'un système à une seule sortie, consiste en un ensemble d'équations de la forme :

$$\mathbf{x}(k) = \mathbf{f} \mathbf{x}(k-1), u(k-1) \quad \text{équation d'état}$$

$$y(k) = g \mathbf{x}(k) \quad \text{équation d'observation}$$

où :

- $k$  est l'instant présent discrétisé,
- $\mathbf{u}(k)$  est le vecteur des variables d'entrée,
- $\mathbf{x}(k)$  est le vecteur des variables d'état,
- $\mathbf{f}$  est une fonction vectorielle,
- $g$  est une fonction scalaire,
- $y(k)$  est la sortie de la représentation d'état du système.



**Figure 7 : Représentation d'état**

Si le vecteur d'état  $\mathbf{x}(k)$  est complètement accessible à la mesure à chaque instant, un réseau de neurones peut identifier les fonctions  $\mathbf{f}$  et  $g$ . Dans le cas contraire, il n'y a en général pas d'unicité de la fonction attendue, l'interprétation des grandeurs d'état ainsi délivrées sera plus difficile à réaliser.

Dans le cadre de la prévision des crues qui motive ce travail, le vecteur des variables d'état pourrait être constitué des débits à différents points du réseau hydrographique, tandis que la sortie serait la hauteur d'eau mesurée à l'aval de ce réseau hydrographique. L'équation d'observation effectuerait, en quelque sorte, une transformation du type de l'inverse de la courbe de tarage.

Lorsque l'on définit comme vecteur d'état le vecteur composé des sorties du processus aux valeurs passées, on parle d'une représentation entrée-sortie (Figure 8), cette dernière peut s'écrire de la façon suivante :

$$y(k) = h \ y(k-1), \dots, y(k-r), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m)$$

où :

- $k$  est l'instant présent discrétisé,
- $y(k)$  est la sortie de cette représentation,
- $h$  est une fonction,
- $\mathbf{u}(k)$  est le vecteur des variables exogènes,

- $m$  et  $r$  sont des entiers positifs.

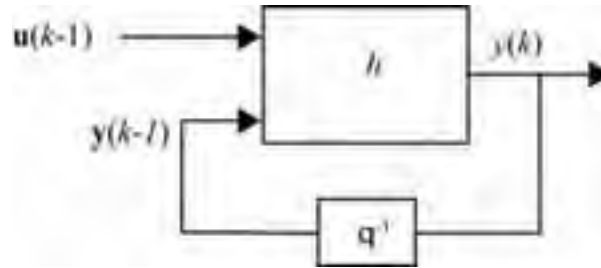


Figure 8: Représentation entrée-sortie

Une fois ces équations générales posées il convient de préciser certaines hypothèses qui vont conduire à la proposition de différents modèles postulés pour le processus ; en particulier on devra choisir si le processus est linéaire ou non linéaire et comment interviennent les perturbations dans le modèle postulé.

#### 1.4.1.b Prise en compte des perturbations dans les modèles prédictifs entrée-sortie

##### ➤ Hypothèse bruit d'état

Intéressons nous tout d'abord au cas où le modèle postulé pour le processus est soumis à des perturbations additives sur son état. Dans ce cas, la relation entrée-sortie du modèle postulé s'écrit :

$$y^p(k+1) = \varphi(y^p(k), \dots, y^p(k-r+1), u(k), \dots, u(k-m+1)) + b(k+1)$$

où :

- $y^p(k+1)$  est la sortie du processus que l'on veut prédire, mesurée à l'instant  $k+1$ ,
- $u(k)$  est le vecteur des variables d'entrée à l'instant présent  $k$ ,
- $b(k+1)$  est le vecteur des perturbations non mesurées à l'instant futur  $k+1$ .

Cette situation est représentée en Figure 9-a ; on remarque que le bruit est représenté par une addition sur la sortie au temps  $k+1$ , mais a influencé également les sorties aux temps précédents.

Cette hypothèse correspond au modèle NARX (Non linéaire Auto-Régressif à entrées exogènes).

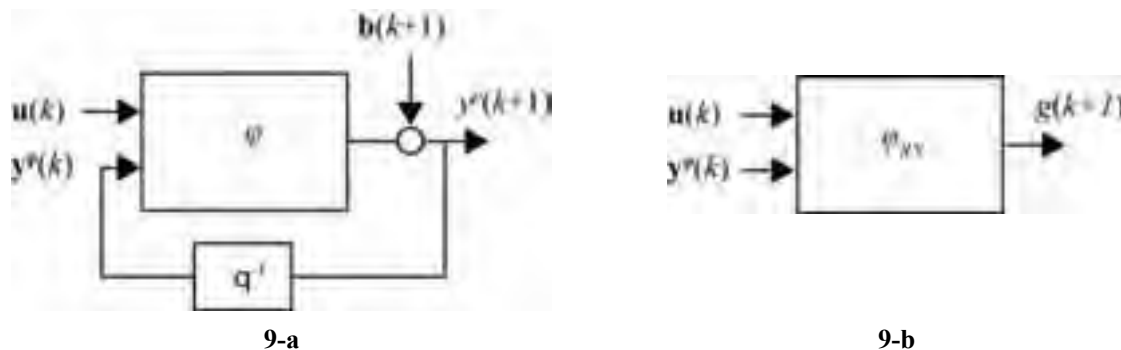


Figure 9 : Modèle postulé (9-a) et prédicteur neuronal (9-b) associés à l'hypothèse *bruit d'état*

Compte tenu de ce modèle postulé, le réseau de neurones devant effectuer la prédiction de la grandeur de sortie du processus réalise la fonction ( $w$  est le vecteur des paramètres du réseau) :

$$g(k+1) = \varphi_{RN} \ y^P(k), \dots, y^P(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-m+1), \mathbf{w}$$

Si la fonction  $\varphi_{RN}$  parvient à identifier parfaitement la fonction  $\varphi$ , c'est-à-dire à réaliser l'identification la meilleure possible, alors on remarque que  $y^P(k) - g(k) = b(k)$ , pour tout  $k$  ; ainsi le modèle a pu représenter tout ce qui est déterministe et son erreur est limitée au bruit d'état non mesurable. Il s'agit donc du meilleur modèle possible.

Il est intéressant de noter que le modèle prédictif dont la sortie est  $g(k)$  n'est pas récurrent, comme illustré en Figure 9-b, même s'il a pour vocation de prédire le comportement d'un processus dynamique ; les méthodes utilisées pour son apprentissage sont donc celles développées pour les réseaux statiques ; comme cet apprentissage est dirigé par les valeurs mesurées sur le processus, on appelle cet apprentissage *l'apprentissage dirigé*.

On peut remarquer que ce prédictif utilise en entrée les grandeurs mesurées du processus : il ne peut être utilisé que pour effectuer une prédiction à *un pas*. Si l'on désire étendre l'horizon de prévision, on peut substituer les sorties estimées du processus aux variables mesurées, mais dans ce cas les performances du prédictif sont dégradées.

Dans la pratique, ce prédictif est celui qui délivre les meilleures prédictions sur les crues, ce qui pourrait laisser penser que les incertitudes sur les mesures de pluies ont plus d'impact sur la modélisation que celles qui sont dues aux mesures de débits ; ces dernières sont prises en considération par le modèle hypothèse suivant.

➤ *Hypothèse bruit de sortie*

Le modèle supposé pour le processus est, dans ce cas, soumis à des perturbations additives sur sa sortie (Figure 10-a) ; ceci se traduit par les relations suivantes :

$$x_p(k+1) = \varphi \ x_p(k), \dots, x_p(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-m+1)$$

$$y^P(k+1) = x_p(k+1) + b(k+1)$$

où

- $x_p(k+1)$  représente la sortie d'état du processus (même si ce dernier est représenté avec la modélisation entrée-sortie, il est décrit par un vecteur d'état composant ses sorties bouclées) au temps de prévision  $k+1$ ,
- $\mathbf{u}(k)$  est le vecteur des variables exogènes à l'instant présent  $k$ ,
- $b(k+1)$  est le bruit de mesure, représenté par une perturbation sur la seule sortie à l'instant futur  $k+1$ .

Le choix de ce modèle postulé sous-entend que l'état du processus est estimé correctement par le modèle et que ce vecteur des sorties non observées est bouclé sur l'entrée du modèle. En revanche, la sortie mesurée étant bruitée on évite de l'utiliser dans l'estimation des valeurs futures.

Le modèle neuronal correspondant à ce processus supposé s'écrit alors ( $\mathbf{w}$  est le vecteur des paramètres du réseau) :

$$g(k+1) = \varphi_{RN} \ g(k), \dots, g(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-m+1), \mathbf{w}$$



Comme précédemment si la fonction  $\varphi_{RN}$  identifie parfaitement la fonction  $\varphi$ , et si le vecteur  $\mathbf{g}$  est initialisé avec des erreurs égales au vecteur des perturbations, alors  $y^p(k+1) - g(k+1) = b(k+1)$  : le prédicteur réalise la meilleure prédiction possible.

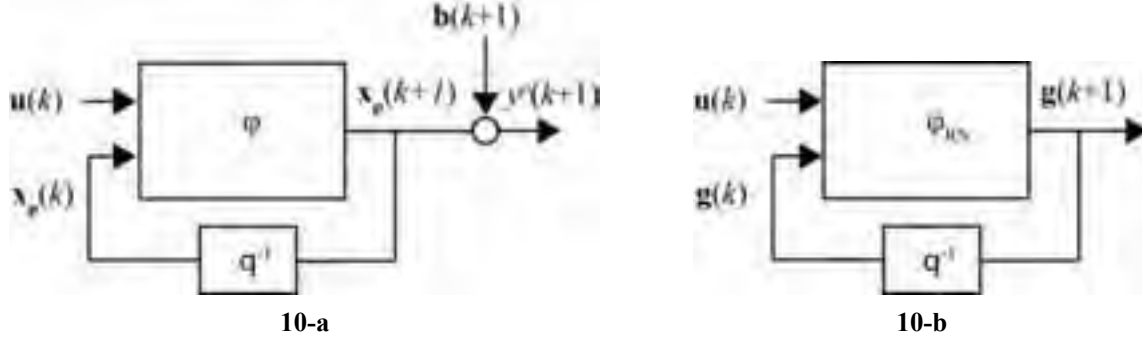


Figure 10 : Modèle postulé (10-a) et prédicteur neuronal (10-b) associé à l'hypothèse *bruit de sortie*

Il apparaît sur la Figure 10-b que le prédicteur ainsi réalisé est récurrent : ses sorties estimées sont appliquées comme variables d'entrée. Il n'est donc pas forcément stable et l'apprentissage doit être mené en prenant en compte les dérivées partielles sur une certaine profondeur temporelle (§I.3.3). L'apprentissage réalisé est de type *semi-dirigé* car seule l'initialisation des copies du prédicteur aux instants précédents est réalisée par les valeurs issues du processus.

Contrairement au prédicteur précédent (NARX), le prédicteur bouclé peut réaliser des prévisions pour un horizon de prévision supérieur à 1. Dans ce cas, il est cependant nécessaire de disposer des valeurs des variables exogènes pour les instants futurs ce qui n'est pas toujours possible pour l'anticipation de phénomènes non contrôlables, en particulier les phénomènes naturels comme les crues ou tremblements de terre. Pour cette raison, et dans ce type d'application, ce prédicteur est également limité à l'anticipation à *un pas*.

➤ *Hypothèse bruit de sortie et bruit d'état*

Le modèle postulé est ici soumis aux deux types de perturbations non mesurables : sur l'état et sur la sortie. Il est décrit par les équations suivantes :

$$x_p(k+1) = \varphi(x_p(k), \dots, x_p(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-m+1), b(k), \dots, b(k-p+1))$$

$$y^p(k+1) = x_p(k+1) + b(k+1)$$

Cette représentation est connue sous le nom de NARMAX (Non linéaire Auto Régressif à Moyenne Ajustée et entrées eXogènes). De la même manière que précédemment, le modèle neuronal doit être récurrent et disposer, de plus, des variables de sorties mesurées du processus. Sous réserve que le modèle neuronal soit stable dans le domaine d'utilisation et que la fonction  $\varphi_{RN}$  du réseau de neurones identifie parfaitement la fonction  $\varphi$ , on peut montrer que ce réseau peut théoriquement modéliser tout ce qui est déterministe dans la relation entrée-sortie, et ne modélise pas le bruit [DREYFUS *et al.* 2008]

## I.4.2 La modélisation pour la prévision

### I.4.2.a Prévision à un horizon supérieur à une période d'échantillonnage

La prévision est l'estimation de la sortie d'un processus physique à un instant postérieur à l'instant présent  $k$ . Le laps de temps séparant l'instant auquel la prévision est souhaitée et l'instant présent  $k$  est l'horizon de prévision  $h_p$ . Pour un modèle discret,  $h_p$  est donc un multiple de la période d'échantillonnage  $T$ .

Rappelons que la prévision se distingue de la simulation qui vise à représenter le fonctionnement du système sans anticipation ; pour la simulation, les entrées, sorties et états sont pris en compte jusqu'aux instants  $k$  pour représenter le fonctionnement du processus à ce même instant  $k$ . Les modèles de simulation sont intéressants pour approfondir la connaissance des processus physiques étudiés ; simuler n'est pas prévoir !

### I.4.2.b Les réseaux de neurones en prévision : illustrations

L'intérêt pratique des réseaux de neurones n'est plus à démontrer. Ils sont largement utilisés en prévision dans différents domaines, tant en sciences de l'ingénieur qu'en sciences humaines ou financières. Ainsi, plutôt que de présenter une vaste liste de toutes les utilisations réalisées (qui ne pourrait prétendre être exhaustive tant elles sont nombreuses), nous préférons illustrer les concepts présentés dans ce chapitre par quelques cas concrets dans le domaine applicatif de cette thèse : l'hydrologie.

La comparaison entre les modèles hypothèses NARX, avec hypothèse de bruit d'état, et le modèle bouclé avec hypothèse de bruit de sortie a été réalisée par [JOHANNET *et al.* 1994, 2008] pour effectuer la prévision à un pas des débits d'une rivière : le Baget (Pyrénées-France). Du fait des incertitudes plus importantes sur l'estimation des entrées exogènes (les précipitations) qui peuvent être vues comme agissant comme une perturbation d'état (Figure 11), le prédicteur neuronal le plus performant s'est avéré être le prédicteur dirigé.

Comme représenté sur la Figure 11, on peut en effet considérer que les erreurs effectuées sur la mesure des entrées extérieures,  $\varepsilon(k)$ , ici déduites des pluies réelles ( $\mathbf{p}^i(k)$ ) moins les pluies mesurées ( $\mathbf{p}^m(k)$ ), agit par l'intermédiaire d'un filtre de fonction  $\psi$  (inconnue) sur l'état du processus alimenté par des pluies exactes.

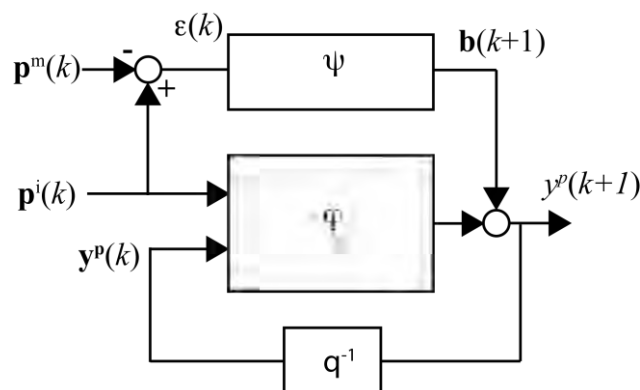


Figure 11 : Représentation des erreurs de mesure de précipitation vues comme une perturbation d'état

Par ailleurs une autre étude, comparant les deux types de modèles postulés, NARX, avec hypothèse de bruit de sortie et NARMAX, a été réalisée dans le domaine de l'hydraulique [ROUSSEL 2001]. Il s'agit de surveiller les niveaux d'eau pluviale dans les réseaux urbains de la Seine-St-Denis à l'aide de capteurs, et de détecter les défaillances des capteurs qui peuvent être sujets à des dérives de l'ordre de 1 cm par jour. En s'intéressant à l'erreur de prédiction, il a pu être mis en évidence que le modèle NARX est inopérant car son hypothèse de bruit d'état ne correspond pas à celle qui se produit réellement. En revanche la dérive des capteurs a pu être mise en évidence en utilisant l'erreur issue du modèle NARMAX : l'erreur de modélisation d'un capteur en bon fonctionnement est à moyenne presque nulle, tandis que l'erreur de modélisation de la hauteur d'eau mesurée par un capteur soumis à une dérive présente une dérive qui peut être détectée à partir d'un certain seuil. Le prédicteur NARMAX a donc bien pu apprendre le comportement du capteur indépendamment de l'erreur de mesure de sortie (la dérive).

Ces deux applications illustrent l'intérêt de choisir le bon modèle hypothèse pour réaliser un prédicteur efficace. En pratique, pour la modélisation de processus physiques naturels ou artificiels, le rôle des perturbations n'est pas toujours bien connu et il est nécessaire d'envisager leurs différents moyens d'action au travers des différents modèles hypothèse qui en découlent.

#### *1.4.2.c Démarche de conception d'un prédicteur neuronal*

Une bonne synthèse de la méthode de conception d'un prédicteur neuronal est présentée dans [DREYFUS *et al.* 2008] ; il nous paraît intéressant de la rappeler synthétiquement dans ce mémoire car elle a servi de canevas à notre étude.

Après avoir vérifié que le processus présentait un fonctionnement non linéaire, la démarche de définition du prédicteur de ce système doit proposer des solutions aux problèmes suivants :

- Prétraitement et sélection des variables pertinentes, tant pour les entrées que pour les sorties (il n'y a pas toujours qu'une sortie possible : par exemple, pour la prévision des crues, on peut choisir la hauteur d'eau ou le débit).
- Choix de la représentation du modèle (état ou entrée-sortie).
- Choix du modèle postulé, concernant non seulement les perturbations non mesurables, mais également l'horizon de prévision souhaité en fonction de la disponibilité des entrées exogènes futures.
- Choix de l'ordre du modèle.
- Sélection de la complexité du modèle (le nombre de couches et de neurones cachés).

Cette démarche, implémentée notamment dans le logiciel NeuroOne de NETRAL S.A.<sup>1</sup>, a permis de résoudre de très nombreux problèmes nécessitant la mise en œuvre d'une méthode d'apprentissage statistique.

---

<sup>1</sup> <http://www.netral.com>

## ***1.5 Conclusion***

Dans ce chapitre, nous avons présenté les définitions et propriétés fondamentales des réseaux de neurones. Leur utilisation est avantageuse en raison de leur propriété d'approximation parcimonieuse de fonction, et ils sont appropriés lorsqu'il s'agit de modéliser des processus statiques ou dynamiques

La présentation des réseaux de neurones nous a amenés à expliciter l'apprentissage sous l'angle statistique. Ce dernier se fait sur un ensemble de couples entrées-sorties connus en nombre restreint afin d'estimer les paramètres du réseau en minimisant l'erreur quadratique moyenne.

Par ailleurs, pour éviter le surajustement du modèle aux exemples d'apprentissage, qui aurait pour conséquence d'augmenter l'erreur de généralisation à des exemples inconnus, les méthodes de régularisation ont été présentées. Ces dernières contraignent l'apprentissage des modèles aussi bien en phase d'estimation des paramètres qu'en phase de sélection de la complexité. Du fait de la disponibilité des exemples en nombre restreint et de la présence de perturbations sur les processus réels, ces méthodes permettent de résoudre au mieux le dilemme biais-variance.

Enfin, la modélisation de processus dynamiques pour la prévision a été présentée ; nous avons introduit la notion d'horizon de prévision, présenté les différents prédicteurs optimaux en fonction des types de perturbations opérant sur le processus physique et ses moyens d'observation. Parmi les applications très nombreuses des réseaux de neurones en prévision, quelques exemples dans le domaine de l'hydrologie ont permis de mettre en évidence la pertinence de la présentation précédente. Ainsi, la prévision des crues dévastatrices cévenoles pourra être abordée dans un cadre méthodologique rigoureux et déjà éprouvé.

## ***1.6 Références bibliographiques du Chapitre I***

- [BARRON 1993] A. R. BARRON. *"Universal Approximation Bounds for Superpositions of a Sigmoidal Function"*. IEEE Transactions on Information Theory IT-39, pp. 930-945, 1993
- [BARTLETT 1997] P. L. BARTLETT. *"For valid generalization, the size of the weights is more important than the size of the network"*. Neural Information Processing Systems, vol 9, Morgan Kaufmann, 1997
- [BENDER 1996] E. BENDER. *"Mathematical Methods in Artificial Intelligence"*. California, IEEE Computer Society Press, 636 p., 1996
- [CYBENKO 1989] G. CYBENKO. *"Approximation by Superpositions of a Sigmoidal Function"*. Mathematics of Control, Signals and Systems, vol 2, pp. 303-314, 1989
- [DREYFUS *et al.* 2008] G. DREYFUS, J.-M. MARTINEZ, M. SAMUELIDES, *et al.* *"Apprentissage statistique"*. EYROLLES, Algorithmes, 417 p., 2008
- [FUNAHASHI 1989] K. FUNAHASHI. *"On the Approximate Realization of Continuous Mappings by Neural Networks"*. Neural Networks, vol 2, pp. 183-192, 1989
- [GALLINARI *et al.* 1999] P. GALLINARI, T. CIBAS. *"Practical complexity control in multilayer perceptrons"*. Signal Processing, vol 74, pp. 29-46, 1999
- [GEMAN *et al.* 1992] S. GEMAN, E. BIENENSTOCK, R. DOURSAT. *"Neural networks and the bias / variance dilemma"*. Neural Computation, vol 4, pp. 1-58, 1992
- [HORNIK *et al.* 1989] K. HORNIK, M. STINCHCOMBE & H. WHITE. *"Multilayer feedforward networks are universal approximators."*. Neural Networks, vol 2, pp. 359-366, 1989
- [HORNIK *et al.* 1990] K. HORNIK, M. STINCHCOMBE, H. WHITE. *"Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks"*. Neural Networks, vol 3, pp. 551-560, 1990
- [HORNIK 1991] K. HORNIK. *"Approximation capabilities of multilayer feedforward networks"*. Neural Networks, vol 4, pp. 251-257, 1991
- [JOHANNET *et al.* 1994] A. JOHANNET, A. MANGIN, D. D'HULST. « Subterranean Water Infiltration Modelling by Neural Networks: Use of Water Source Flow »- . Proc. of ICANN; pp1033-1036. M. Marinaro and P.G. Morasso eds, Springer (1994).
- [JOHANNET *et al.* 2008] A. JOHANNET, A. MANGIN, B. VAYSSADE. *"Modélisation d'un système karstique par réseaux de neurones : Simulation des débits du karst du Baget - France"*. Collection EDYTEM – Cahiers de Géographie – n° 7 - 2008
- [MCKAY 1992] D. J. C. MCKAY. *"A practical Bayesian framework for backpropagation networks"*. Neural Computation, vol 4, pp. 448-472, 1992
- [MEYER 1990] Y. MEYER. *"Ondelettes et Opérateurs I : Ondelettes."*. Editions Hermann, 1990

- [NERRAND *et al.* 1993] O. NERRAND, P. ROUSSEL-RAGOT, L. PERSONNAZ, G. DREYFUS, S. MARCOS. *"Neural networks and non-linear adaptive filtering : unifying concepts and new algorithms"*. Neural Computation, vol 5, pp. 165-197, 1993
- [POGGIO *et al.* 1985] T. POGGIO, V. TORRE, C. KOCH. *"Computational vision and regularization theory"*. Nature, vol 317, pp. 314-319, 1985
- [POWELL 1985] M. J. D. POWELL. *"Radial Basis Functions for Multi-variable Interpolation : A review"*. IMA Conference on Algorithms for the Approximation of Functions and Data, RMCS Shrivenham, UK, 1985
- [PRESS 1992] J.W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, B. P. FLANNERY. «Numerical recipes in C». Cambridge University Press 1992
- [ROUSSEL *et al.* 2001] P. ROUSSEL, F. MONCET, B. BARRIEU, A. VIOLA. *"Modélisation d'un processus dynamique à l'aide de réseaux de neurones bouclés. Application à la modélisation de la relation pluie-hauteur d'eau dans un réseau d'assainissement et à la détection de défaillance de capteur"*. Innovative Technologies in urban drainage, 1, 919-926, G.R.A.I.E., 2001.
- [RUMELHART *et al.* 1986] D. E. RUMELHART, G. E. HINTON & R. J. WILLIAMS. *"Learning Internal Representations by error backpropagation"*. Parallel Distributed Processing : Explorations in the Microstructure of Cognition, pp. 318-362, MIT Press, 1986
- [SAARINEN *et al.* 1993] S. SAARINEN, R. BRAMLEY, G. CYBENKO. *"Ill-conditioning in neural network training problems"*. SIAM J. Sci. Stat. Comp., vol 14, pp. 693-714, 1993
- [SJÖBERG *et al.* 1995] J. SJÖBERG, Q. ZHANG, L. LJUNG *et al.* *"Nonlinear black-box modeling in system identification : a unified overview"*. Automatica, vol 31, pp. 1691-1724, 1995
- [STONE 1974] M. STONE. *"Cross-validatory choice and assessment of statistical predictions"*. Journal of the Royal Statistical Society, B36, pp. 111-147, 1974
- [TIKHONOV *et al.* 1977] A. N. TIKHONOV, V. Y. ARSENIN. *"Solutions of Ill-Posed Problems"*. Winston, 1977
- [VAPNIK 1995] V. N. VAPNIK. *"The Nature of Statistical Learning Theory"*. Springer, 1995
- [YACOUB *et al.*] M. YACOUB, Y. BENNANI « Une mesure de pertinence pour la sélection de variables dans les perceptrons multicouches ». Revue d'intelligence artificielle, vol.15, pp. 393-410, 2001
- [ZHOU *et al.* 1998] G. ZHOU, J. SI. *"A systematic and effective supervised learning mechanism based on jacobian rank deficiency"*. Neural Computation, vol 10, pp. 1031-1045, 1998



## II Prévision des crues éclair du Gardon d'Anduze

### II.1 Introduction

Les bassins versants montagneux situés sur le pourtour de la Méditerranée sont sujets à des crues éclair caractéristiques et extrêmement violentes. Parmi celles-ci, les *Gardonnades* sont l'objet de notre étude. Plus particulièrement, nous présentons en première partie de ce chapitre le bassin versant d'Anduze et le projet de modélisation des crues qui a motivé cette étude : le Bassin Versant Numérique Expérimental des Gardons. Nous verrons ensuite que plusieurs études hydrologiques se focalisent sur ce bassin versant et nous nous attacherons à présenter ces études afin de mieux comprendre les phénomènes en jeu, ce qui nous permettra de mieux adapter la démarche de conception du réseau de neurones. Puis nous nous intéresserons au cahier des charges et à la possibilité de le mettre en œuvre avec les données disponibles. Seront ainsi envisagés tour à tour le choix de l'échantillonnage, la détermination du type de capteur de pluie, et le choix de la grandeur physique de sortie. Nous verrons que malgré de nombreuses études et beaucoup de mesures disponibles, la difficulté de la récolte de données durant les épisodes de pluies intenses conduit à des erreurs de mesures et à des erreurs d'estimation considérables, laissant augurer de réelles difficultés pour parvenir à la prévision opérationnelle de ces crues.

### II.2 Les Gardonnades

Les crues éclair sont engendrées par des pluies convectives concentrées sur une zone localisée pendant plusieurs heures ; si l'étendue des systèmes convectifs de méso-échelles<sup>2</sup> peut couvrir de larges zones des Cévennes, les ruissellements à l'origine de ces crues se produisent sur des bassins versants de taille petite ou moyenne (10-100 km<sup>2</sup>). Elles sont souvent difficilement prévisibles, de temps de montée rapide et de débit spécifique relativement important<sup>3</sup>. L'impact dévastateur de ces crues est d'autant plus grand que le bassin versant est anthropisé. Ainsi, même en France où les services de l'état ont accumulé de nombreuses études hydrologiques depuis plus d'un siècle, on peut trouver des bassins versants, qui, du fait de leur petite taille et de l'extrême difficulté de réaliser des mesures durant ces événements cataclysmiques, n'ont pu être étudiés avec toute la qualité nécessaire. C'est le cas des bassins versants cévenols, et en particulier de celui du Gardon. Chacun se souvient d'avoir un jour entendu au journal télévisé l'annonce d'orages sur les « versants sud des Cévennes » devant avoir pour conséquence des crues, pour l'instant imprévisibles, tant pour leur localisation précise que pour leur débit de pointe ; ces crues sont les « Gardonnades ».

D'une superficie de 543 km<sup>2</sup> (Figure 12), le bassin versant du Gardon d'Anduze à Anduze est de petite taille ; son altitude varie de 111 m à 1366 m. Ses versants d'écoulement ont de fortes pentes (environ 30 % [AYRAL 2005]), ce qui favorise des temps de concentration (laps de

---

<sup>2</sup> Méso-échelle signifie que la dimension spatiale prise en considération est intermédiaire entre la dimension globale (planétaire) et la dimension des nuages individuels dits de micro-échelle. Elle désigne des systèmes dont l'extension va de 2 km à 2000 km.

<sup>3</sup> Le débit spécifique est le débit rapporté à la surface du bassin versant. Il s'exprime en m<sup>3</sup>/(s.km<sup>2</sup>).



temps que met l'eau d'une averse pour parvenir au réseau hydrographique) très courts. Le temps de concentration du bassin versant d'Anduze est estimé à environ 4 heures [MARCHANDISE 2007], et le temps de montée de 3 à 5h [BOUVIER 2006].



**Figure 12 : Le bassin versant d'Anduze et sa situation en France**

La durée très courte de concentration des eaux justifie l'appellation de rapide ou d'*éclair* pour décrire ces crues. On comprend aisément que le risque induit par ces événements soit difficile à gérer et nécessite le recours à des modélisations mathématiques afin de disposer de prévisions hydrologiques mais surtout météorologiques fiables.

La réalisation de modèles de prévision des crues éclair du Gardon à Anduze est l'objet du projet *Bassin Versant Numérique Expérimental* (ou BVNE) *Gardon*, projet organisé à l'initiative du Service Central d'Hydrométéorologie et d'Appui aux Inondations (ou SCHAPI), service décentralisé du Ministère chargé de l'Ecologie, en 2006. Ce projet s'appuie sur le Service de Prévision des Crues (ou SPC) *Grand Delta*, ainsi que sur des laboratoires de

recherche développant et adaptant leurs modèles hydrologiques afin d'effectuer, en temps réel, une prévision des débits à Anduze.

## **II.3 Etudes hydrologiques sur le bassin versant d'Anduze**

### **II.3.1 La modélisation hydrologique**

De l'accord de tous les hydrologues, le fonctionnement hydrologique est très complexe et les phénomènes physiques à la base du ruissellement sont incomplètement connus. Néanmoins, afin de mieux comprendre la difficulté de la prévision des crues éclair, et de présenter l'apport que peuvent avoir les réseaux de neurones dans ce contexte, nous nous attacherons à présenter les principales hypothèses explicatives du ruissellement ainsi que les principaux modèles utilisés. La présentation de quelques modèles appliqués au Gardon d'Anduze permettra d'illustrer notre propos.

Tout modèle hydrologique convertit l'eau tombée sur un versant en débit à l'exutoire de ce bassin. La pluie est donc la première variable prise en considération par le modèle. Si le rôle de l'hétérogénéité spatiale des précipitations est encore discuté, il apparaît néanmoins que cette hétérogénéité est très importante dans la genèse des crues éclair et doit être prise en considération [BORREL 2004], [LE LAY *et al.* 2007], [SAULNIER *et al.* 2009]. Par ailleurs, une fonction non négligeable du modèle est de répartir les pluies tombées en pluies efficaces (celles qui vont contribuer au ruissellement constitutif de la crue), en pluies infiltrées (celles qui vont participer à l'humidification du bassin versant), en pluies évaporées, transpirées (consommées par la végétation puis éventuellement délivrées à l'atmosphère) ou en pluies interceptées (en mouillant la végétation). En ce qui concerne les crues éclair, du fait même de la grande quantité d'eau précipitée en un temps très court (ordre de grandeur 24 h), il est raisonnable de penser que la définition précise de la quantité d'eau évaporée n'est pas le point le plus critique du modèle.

En revanche il est intéressant de noter que l'absence de prévision de pluies aux échelles temporelles et spatiales de travail (moins d'une heure et quelques dizaines de km<sup>2</sup>) est une réelle difficulté pour les prévisions hydrologiques<sup>4</sup>. Classiquement cette difficulté est abordée en considérant que les pluies passées observées sont prolongées durant l'horizon de prévision déterminé, ou bien que la pluie s'arrête durant cet horizon temporel (l'hypothèse de pluie nulle est prise comme une référence objective, mais elle constitue l'hypothèse météorologique la plus optimiste). Dans les deux cas, ces hypothèses conduisent à de lourdes erreurs ou à des insuffisances pour la prévision des crues éclairs, au cours desquelles les pluies peuvent varier de manière tout à fait imprévisible tant localement que spatialement. On est donc conduit à définir deux types de modèles : les modèles de simulation qui s'attachent à estimer les débits actuels à partir des pluies passées - ces modèles auront tout leur intérêt en prévision lorsque l'on disposera de prévision de pluies -, et les modèles de prévision qui s'attachent à estimer les débits futurs avec l'une ou l'autre des deux hypothèses de pluies adoptées, ou sans aucune hypothèse de pluie. Sur ce dernier point, nous montrerons que les réseaux de neurones apportent une autre vision de la prévision par le fait même qu'ils ne revendiquent pas un

---

<sup>4</sup> Cette lacune est en train d'être comblée par l'arrivée du modèle AROME de METEO FRANCE [BOUTTIER 2003]. Ce dernier n'est cependant pas encore opérationnel aux échelles temporelles et spatiales nécessaires pour cette étude.

fonctionnement physique. Les débits futurs pourront être estimés en fonction des pluies passées, sans hypothèse sur les pluies futures, comme cela se fait pour les séries chronologiques.

Une fois la pluie prise en considération, les modèles hydrologiques définissent ce que l'on appelle classiquement la fonction de production et la fonction de transfert<sup>5</sup>. Grossièrement, la fonction de production est celle qui détermine l'eau qui ruisselle jusqu'au cours d'eau en fonction des précipitations et des variables d'état du bassin. La fonction de transfert s'intéresse à la propagation de l'eau dans le cours d'eau. La propagation de l'eau dans le cours d'eau voit schématiquement sa vitesse décroître d'amont en aval, au fur et à mesure que les pentes diminuent, et voit corrélativement ses hauteurs augmenter vers l'aval. Historiquement les fonctions de production et de transfert étaient indépendantes et étaient estimées en tant que telles. Avec l'avènement des modélisations hydrologiques à base physique, qui prennent en considération l'hétérogénéité spatiale (modèles distribués), ces deux fonctions perdent un peu de leur sens puisqu'elles sont réduites spatialement à des mailles élémentaires et donc couplées l'une à l'autre. De plus, elles sont de plus en plus dépendantes des hypothèses physiques postulées sur les processus hydrologiques et hydrauliques mis en jeu dans les modèles.

La fonction de production est fondamentale dans la description des petits bassins pour lesquels la propagation peut avoir un rôle moindre du fait de leur petite taille. Actuellement, deux hypothèses explicatives de la genèse du ruissellement s'opposent : il s'agit des ruissellements hortonien et hewlettien.

Le ruissellement hortonien [HORTON 1933] se produit localement lorsque l'intensité des pluies dépasse la capacité d'infiltration des sols. Une fois engendré sur les versants, le ruissellement hortonien peut se ré-infiltrer en circulant à la surface ou rejoindre le cours d'eau en contrebas, favorisant un accroissement du débit. Dans ce fonctionnement, l'intensité des pluies est la variable prépondérante dans l'explication du volume ruisselé.

Le ruissellement hewlettien [HEWLETT 1961] suppose au contraire que les sols ont une capacité d'infiltration importante tant qu'ils ne sont pas saturés en eau ; c'est cette saturation qui les rend imperméables. L'estimation du ruissellement total se fait donc au moyen de la détermination des zones saturées en eau, l'extension de ces dernières dépendant de la nature des sols et de la topographie du versant. C'est pourquoi l'hypothèse de Hewlett est aussi appelée *méthode des aires contributives*<sup>6</sup>. Dans ce fonctionnement, l'état d'humidité du bassin versant est la variable prépondérante pour expliquer la genèse des crues.

---

<sup>5</sup> La fonction de transfert ne doit pas être confondue avec la représentation d'un système linéaire dans le formalisme de Laplace ; il s'agit du transfert physique de l'eau.

<sup>6</sup> En toute rigueur le ruissellement est créé sur une surface saturée. Ce ruissellement peut se ré-infiltrer lors de son parcours sur le versant et donc finalement ne pas contribuer au débit. Les surfaces contributives sont donc en toute rigueur les surfaces saturées dont le ruissellement contribuera effectivement au débit. Par abus de langage on confond souvent les deux termes "saturées" et "contributives". Sur les bassins cévenols on peut vérifier que cette confusion est sans conséquence [SAULNIER 1997].

Si les deux hypothèses précédentes font appel à une représentation physique de l'écoulement de l'eau dans le versant<sup>7</sup>, il existe cependant d'autres modèles, fondés sur des concepts intuitifs tels que le réservoir, dont on sait qu'ils sont physiquement inexacts, mais dont la détermination des paramètres par calage peut donner lieu à des courbes de simulation proches de la réalité. Les modèles à réservoir sont appelés *conceptuels*, même si cette appellation est ambiguë puisque les modèles hewlettien et hortonien font aussi appel à des concepts, ces derniers étant d'ordre physique. Un des premiers modèles à réservoir simulés en France a été conçu par le Centre d'Etude du Machinisme Agricole, du Génie Rural, des Eaux et des Forêts (CEMAGREF) [MICHEL, 1983], il s'agit du modèle GR. Cette première version de modèle à réservoir ne comportait qu'un seul paramètre à ajuster. La robustesse en extrapolation des modèles conceptuels est souvent discutée par rapport à celle des modélisations physiques.

Partant de ces principes, plusieurs modèles hydrologiques sont en cours d'étude sur le bassin versant d'Anduze, dans le cadre du projet BVNE. Nous présentons dans ce qui suit les démarches de modélisation qui ont été utilisées pour effectuer la prévision des crues du Gardon d'Anduze. Ces démarches sont intéressantes à considérer si l'on veut mieux comprendre l'apport de l'apprentissage statistique à ce problème.

### **II.3.2 Application de modèles hydrologiques au Gardon d'Anduze**

#### *II.3.2.a TOPMODEL*

TOPMODEL (*TOPography based hydrological MODEL*) est un modèle à base physique fondé sur l'hypothèse des aires contributives ; il est appliqué aux bassins versants des Cévennes-Vivarais dans le cadre du projet BVNE et du projet européen PREVIEW par G. M. Saulnier et M. Le Lay. TOPMODEL calcule deux contributions aux débits (selon une période qui varie de 15 minutes à 1 heure en fonction de la taille du bassin) : une partie exfiltrée du sol et le ruissellement sur les surfaces saturées. Comme indiqué dans l'introduction, l'hypothèse des aires contributives peut être appliquée correctement si l'on dispose d'informations spatialisées sur le bassin versant considéré (Modèle Numérique de Terrain), permettant de déterminer quelles sont les aires qui sont saturées au cours du temps, et quels ruissellements ces aires produisent. Ce comportement du bassin versant est confirmé par les travaux de [LE LAY *et al.* 2007] qui a étudié les débits à 19 points de mesures correspondant à autant de sous bassins versants, dont certains sont emboîtés, dans les bassins versants du Gard à Remoulins, de la Cèze et de l'Ardèche, ces deux dernières rivières étant voisines du bassin versant du Gardon. M. Le Lay a utilisé une version du modèle hydrologique travaillant à une résolution spatiale de 20 m (la taille du pixel du modèle numérique de terrain) et une résolution temporelle horaire. Il a ensuite adopté une démarche visant à prendre en compte les

---

<sup>7</sup> De plus en plus la dichotomie *modèle physique/modèle conceptuel* est remise en cause car les lois physiques utilisées dans les modèles à concepts physiques sont utilisées parfois hors de leur domaine de validité, et leurs paramètres ne résistent pas à un changement d'échelle. De ce fait ces paramètres ne sont pas obtenus par mesures physiques mais par calage, ce qui leur ôte toute validité physique. [LE LAY 2006] propose une présentation unifiée qui permet de distinguer au sein d'un même modèle certains sous-processus opérant à l'échelle adéquate dont l'équation différentielle peut être résolue et d'autres opérant à des échelles plus importantes qui sont paramétrés (comme l'évaporation).

différences spatiales des variables exogènes du modèle et/ou de ses paramètres, sur les mailles de dimension d'environ 50 km<sup>2</sup>. Au début, une description uniforme de toutes les mailles pour toutes les variables est réalisée, puis les variables sont raffinées spatialement dans l'ordre suivant : raffinement d'un index topographique moyen sur chaque maille (description de la pente), puis spatialisation des précipitations, puis de la teneur en eau des sols initiale, puis des paramètres hydrologiques, et enfin de la rapidité du transfert.

La qualité des résultats de simulation est ensuite observée sur l'événement pluvieux de 2002, le plus intense de ces dernières années, qui sera présenté dans la section II.5.3.b. Cet événement sert ainsi de discriminant pour déterminer les informations les plus utiles pour caler le modèle efficacement. De manière logique, on observe qu'un premier saut qualitatif est obtenu par la spatialisation des pluies ; un second saut est obtenu en distribuant la teneur en eau des sols (et donc la saturation future de ces zones) ; les autres améliorations sont moindres. Nous avons précédemment fait l'hypothèse que l'évaporation pouvait être négligée (II.3.1) pendant la crue. Ce dernier résultat n'est pas contradictoire avec notre simplification. L'évaporation a un rôle sur la crue car elle conditionne la répartition spatiale des humidités dans les sols en début d'épisode cévenol, mais son amplitude pendant la crue n'a que peu d'influence compte tenu de l'ampleur des intensités de pluies convectives.

L'analyse de ce travail permet de présenter les qualités de ce modèle : d'une part, il permet d'obtenir le débit en tout point du réseau hydrographique, même si ce dernier ne peut être évalué que sur certaines stations de mesures ; cette propriété est particulièrement intéressante pour étudier le risque d'inondation sur un point particulier du réseau hydrographique pour lequel il n'y a pas de station de mesure, mais où se trouvent des enjeux humains ou matériels. D'autre part il permet de mieux comprendre le fonctionnement physique du bassin versant.

Ce travail confirme le fait que la prise en considération de l'hétérogénéité spatiale des pluies est extrêmement importante. Ce résultat est confirmé par [SAULNIER *et al.* 2009] où les auteurs proposent une méthodologie (illustrée sur les Cévennes) permettant de distinguer l'impact relatif de la variabilité statistique des intensités des pluies convectives et l'impact relatif de leur localisation géographique sur la genèse des crues cévenoles.

Même si le modèle démontre la sensibilité de la réponse hydrologique à la distribution spatiale des précipitations, il peut fonctionner avec des pluies uniformes. Toutefois il fait partie de la famille des modèles que nous avons qualifiés de "modèles de simulation" (II.3.1). En prévision il sera donc dépendant de la qualité des prévisions de pluies et ne pourra donc donner de résultats robustes en l'absence de celles-ci au delà du temps de concentration du bassin versant.

### II.3.2.b MARINE

MARINE (*Modélisation de l'Anticipation du Ruissellement et des Inondations pour des événements Extrêmes*) est un modèle pluie-débit conçu à L'IMFT (Institut de Mécanique des Fluides de Toulouse), spécifiquement pour la simulation et la prévision des crues sur les petits et moyens bassins [BORREL 2004]. Il est également appliqué dans le cadre du projet BVNE Gardon par D. Dartus et H. Bessière [BESSIERE *et al.* 2007]. C'est un modèle à base physique qui peut reproduire des phénomènes opérant à différentes échelles grâce une spatialisation fine des variables d'intérêt : modèle numérique de terrain (images issues du satellite SPOT, pouvant avoir des pixels de taille 20 m), images des précipitations issues des radars de Météo France (pixels de 1 km<sup>2</sup>), couverture du sol et de la végétation, description de

la rivière. La course à la complexité consistant à multiplier les variables descriptives spatialisées est cependant évitée en choisissant les phénomènes physiques adéquats en fonction de l'échelle de description considérée, et en calibrant certaines des variations spatiales non pas individuellement, mais au moyen d'un scalaire multiplicatif global.

À l'opposé de TOPMODEL, l'hypothèse gouvernant l'infiltration est que le sol se sature par refus d'infiltration (type de Horton). L'eau ruisselle puis est ensuite propagée dans le réseau hydrographique.

En fonction du processus physique représenté, certains paramètres sont initialisés à partir des connaissances physiques (géologiques pour les types de sols et la profondeur des sols), ou bien calés. Les derniers développements de MARINE intègrent une procédure d'adaptation des paramètres, ajustés en fonction de l'événement en cours (ou assimilation de données) ; ils permettent d'obtenir d'excellents résultats de simulation [BESSIERE *et al.* 2007].

Étant, comme TOPMODEL, un modèle à base physique, MARINE a également l'avantage de pouvoir délivrer une simulation des débits en tout point du réseau hydrographique, mais souffre des mêmes difficultés à effectuer la prévision des crues en l'absence de prévision des précipitations spatialisées.

Comme nous l'avons souligné dans l'introduction, il est intéressant de constater que les deux modèles, fondés sur des hypothèses de genèse du ruissellement opposées, conduisent néanmoins à des simulations de qualités équivalentes. Si ceci n'est pas très surprenant, on peut de ce fait considérer *a contrario* que ce que ces modèles ont en commun a une validité mieux établie. Il s'agit de l'importance de la prise en considération des prévisions spatialisées des précipitations et de la teneur en eau des sols initiale.

### II.3.2.c SCS

SCS (*Soil Conservation Service*) est également un modèle pluie-débit appliqué dans le cadre du projet BVNE par C. Bouvier à l'Université de Montpellier [BOUVIER *et al.* 2006]. Il s'agit d'un modèle conceptuel qui utilise la notion de réservoir pour représenter le processus d'infiltration. Le bassin versant est vu comme un réservoir dont la vidange se fait avec un débit proportionnel à la hauteur d'eau dans le réservoir. Ce système est un système linéaire intégrateur du premier ordre dont la constante de temps « représente » la conductivité hydraulique. L'interprétation de son fonctionnement est que la pluie remplit un réservoir « sol » dont le niveau varie en fonction de la différence entre la pluie et l'écoulement. Ce réservoir est considéré comme vide au début de chaque événement et peut se remplir grâce à l'extrême transmissivité des sols : ces derniers peuvent en effet absorber jusqu'à 200 ou 300 mm de précipitation par heure avant d'être saturés [BOUVIER *et al.* 2006]. Le modèle est donc calé avec uniquement deux paramètres à ajuster (la transmissivité et la capacité du réservoir) qui ne sont pas spatialement variables ; en revanche, les pluies appliquées sont spatialisées. Ces dernières sont appliquées à l'échelle du versant, ce qui permet également d'estimer le débit en tout point du réseau hydrographique (comme TOPMODEL et MARINE).

Le comportement de SCS en simulation sur le débit à Anduze, est d'une qualité un peu inférieure à celles présentées par les deux modèles précédents, mais on ne peut qu'être satisfait d'obtenir des hydrogrammes réalistes avec un modèle aussi simple.

En revanche la simplicité même du modèle et le fait que ses paramètres soient ajustés sur des valeurs moyennes régionales permet de penser qu'il pourrait être appliqué en généralisation à

d'autres bassins versants voisins ; en particulier sur des bassins versants non jaugés. Cette propriété reste à être évaluée, en particulier par rapport aux performances de TOPMODEL et MARINE appliqués dans des circonstances similaires.

Cependant, compte tenu des hydrogrammes observés en simulation, la prévision des crues éclair ne pourra être réalisée par SCS de manière cohérente qu'avec une prévision des précipitations spatialisées.

### **II.3.3 Synthèse**

La présentation des modèles hydrologiques appliqués aux crues éclair du bassin versant d'Anduze a mis en évidence l'extrême difficulté de prévoir ces événements. Si les modèles de simulation arrivent à bien reproduire les événements, en particulier les plus intenses, en revanche, ils sont pour l'instant dans l'incapacité de prévoir correctement l'hydrogramme d'un événement très intense au delà du temps de concentration du bassin en l'absence de la variable qui est essentielle pour tous les modèles : la prévision des précipitations, tant temporelle que spatiale. Les hypothèses de pluie nulle ou de pluie constante, habituellement utilisées pour effectuer des prévisions de débits sur des bassins beaucoup plus lents, sont, dans ce cas, inadaptées ou insuffisantes.

L'application de ces modèles ne permet pas de décider quel est le type d'infiltration qui opère réellement sur les versants cévenols, mais pour chacun d'eux, les hypothèses forment un tout cohérent avec les variables qui paraissent pertinentes ou indispensables. Parmi les trois variables qui se dégagent de ces études, on peut citer en premier lieu les précipitations spatialisées, puis, selon les modèles, la capacité d'infiltration ou la teneur en eau initiale des sols. La pertinence de ces variables devra donc être évaluée lorsque nous mettrons en œuvre les modèles à apprentissage statistique.

## **II.4 Critères de qualité**

Afin d'évaluer correctement la qualité des modèles proposés, il convient de définir les critères qui seront utilisés. Pour cela, il faut tenir compte de l'utilisateur des prévisions : le prévisionniste a besoin d'une bonne anticipation de la montée du pic et de la synchronisation entre le pic observé et le pic estimé. Mais ceci n'est pas suffisant ; on peut trouver des réponses qui passent exactement par le pic de crue observé et qui délivrent par ailleurs des pointes de crues fictives, susceptibles de déclencher de fausses alertes dont on sait qu'elles ont un effet très négatif sur le public et son comportement futur.

Ainsi, l'évaluation objective des prévisions ne peut être réalisée par un seul critère et devra en faire intervenir plusieurs. Pour ce faire, nous avons choisi deux critères parmi la panoplie classiquement utilisée et nous en avons défini un supplémentaire, appelé critère de hauteur. Ces critères sont présentés dans la section qui suit.

### **II.4.1 Coefficient de détermination, ou critère de Nash**

Le critère de Nash [NASH *et al.* 1970] représente le coefficient de détermination  $R^2$  qui a pour expression :

$$R^2 = 1 - \frac{\sum_{k=1}^n (y^p(k) - y(k))^2}{\sum_{k=1}^n (y^p(k) - \overline{y^p})^2}$$

où :

- $y(k)$  est la sortie estimée par le modèle,
- $y^p(k)$  est la sortie mesurée du processus,
- $n$  est le nombre de couples entrée-sortie sur lesquels porte le calcul du critère,
- $\overline{y^p}$  est la moyenne des  $n$  observations.

Ce critère quantifie la proportion de la variance totale qui est expliquée par la prévision. En hydrologie, il est très utilisé car il permet de comparer les valeurs du critère calculées sur différents événements pluvieux d'intensités très différentes, ou sur différents bassins versants de débits moyens également très différents.

On remarque que maximiser le  $R^2$  revient à minimiser l'erreur quadratique moyenne des sorties estimées. Ceci signifie que ce critère est celui qui est minimisé par l'apprentissage des réseaux de neurones lorsque la fonction de coût est la simple fonction quadratique. Ce critère s'applique donc logiquement pour évaluer la qualité de l'apprentissage.

La valeur maximale que peut prendre le critère de Nash est égale à 1. Ceci traduit une adéquation parfaite entre la courbe prévue et la courbe observée. Si le critère de Nash est nul, alors ceci traduit que le modèle est aussi bon que celui qui se contenterait de prédire la moyenne des valeurs observées. Lorsqu'il est calculé sur un ensemble de test, différent de l'ensemble d'apprentissage, le  $R^2$  peut devenir négatif, ce qui signifie que le modèle est inadéquat, et incapable de prédire même la moyenne des valeurs observées.

Dans les applications en hydrologie, selon le type de rivière étudié, on peut considérer que des critères de Nash supérieurs à 0,6-0,7 sont acceptables et des critères supérieurs à 0,9 sont bons ou excellents. Pour la prévision des crues éclair, un critère de Nash supérieur à 0,8 est déjà considéré comme bon.

Néanmoins, dans le contexte de la simulation des crues, il convient de considérer ce critère avec un œil critique ; en effet, comme il est calculé sur toute la longueur de la crue, il prend en considération autant la montée du pic que sa décroissance, et peut avoir de bonnes valeurs même si l'hydrogramme estimé est très éloigné de l'hydrogramme observé, en termes de qualité d'ajustement du pic.

De plus, cette fois dans le contexte de la prévision, en particulier si la période d'échantillonnage est faible devant le temps de montée de l'événement, on peut obtenir un excellent  $R^2$  en proposant comme prévision la courbe observée décalée de quelques périodes : le critère de  $R^2$  est excellent mais le bénéfice en prévision est nul. Pour traiter ce problème, le critère de persistance, présenté ci-dessous, a été défini.

#### II.4.2 Critère de persistance

Le critère de persistance [KITADINIS *et al.* 1980] calcule la somme des carrés des différences entre les sorties prévues  $y(k+h_p)$  (rappelons que  $h_p$  est l'horizon de prévision) et



les sorties observées  $y^p(k+h_p)$ , rapporté à la somme des carrés des différences entre les valeurs observées aux instants  $k$  et  $k+h_p$ . L'expression du critère de persistance  $C_p$  est la suivante :

$$C_p = 1 - \frac{\sum_{k=1}^n y^p(k+h_p) - y(k+h_p)^2}{\sum_{k=1}^n y^p(k+h_p) - y^p(k)^2}$$

Le critère de persistance vaut 1 si la prévision est parfaite. Si la prévision effectuée est juste la *prévision naïve*, c'est-à-dire la projection qui consiste à proposer que la grandeur d'intérêt aura la même valeur à l'instant  $k+h_p$  qu'à l'instant  $k$ , le critère vaut 0. Si le critère de persistance est négatif, la prévision est encore plus médiocre que la prévision naïve ; il va sans dire que ce dernier cas est particulièrement mauvais : le modèle n'apporte aucune information.

Ce critère, appliqué sur des prévisions, permet donc de comparer la performance du modèle à celle qui serait obtenue par prévision naïve. Cette comparaison est opportune car la prévision naïve peut parfois conduire à des critères de Nash élevés, en particulier si la période d'échantillonnage est faible par rapport à la dynamique du système. A titre d'illustration, les critères de Nash de la prévision naïve de l'événement de 2002 sont présentés en Annexe F, on peut noter que, pour les horizons de prévision courts, ils sont extrêmement élevés (0,95 pour  $\frac{1}{2}$  h).

### II.4.3 Critère de hauteur

Pour avoir une évaluation de la qualité des prévisions sur des portions plus spécifiques de l'événement, telles que la pointe, le franchissement des seuils d'alerte ou la montée de crue, nous avons proposé de restreindre le calcul du critère de qualité à certains segments de l'hydrogramme.

Ainsi, pour évaluer la synchronisation entre les pics prévus et les pics observés, nous proposons un critère opérant essentiellement dans la phase ascendante de la crue, jusqu'à son maximum. Compte tenu des temps de montée des crues éclair et de l'échantillonnage choisi, trois périodes d'échantillonnage avant la pointe de crue jusqu'à une période d'échantillonnage après la pointe de crue sont choisies pour calculer ce critère appelé  $C_h$ . Pour chaque instant sélectionné, on calcule un rapport inférieur à 1 : la plus petite des valeurs observées ou estimées divisée par la plus grande de ces deux valeurs. Le critère est la moyenne de ces rapports.

Pour les événements comportant plusieurs pics, les différentes plages temporelles relatives aux différents pics sont prises en considération par une moyenne des rapports obtenus.

Ainsi, le calcul du critère de hauteur se fait comme suit :

$$C_h = \frac{1}{n_{pics}} \sum_{j=1}^{n_{pics}} \left( \frac{1}{5} \sum_{k_j-3}^{k_j+1} \frac{\min(y^p(j_2), y(j_2))}{\max(y^p(j_2), y(j_2))} \right) \quad (1)$$

où :

- $n_{pics}$  représente le nombre de pointes de crue contenues dans l'événement,
- $k_j$  est l'instant de la pointe de crue  $j$ .

De même que pour les autres critères, plus  $C_h$  est proche de 1, meilleure est l'estimation des pics. Ce critère peut être négatif si le signal estimé est négatif, ce qui n'a pas de sens en hydrologie, mais n'est pas impossible lorsque le modèle effectue une prévision très mauvaise.

Remarquons que le critère de hauteur ne donne pas d'information sur les faux pics de crue c'est-à-dire ceux qui sont prévus par le modèle et ne sont pas observés.

## **II.5 Projet BVNE Gardon**

Le projet BVNE Gardon a été initié en 2006 par le SCHAPI, qui était alors Service Central du MEDAD<sup>8</sup>. Son objet est de réunir plusieurs équipes travaillant sur la prévision des crues éclair et de comparer les prévisions issues de leurs modèles calés avec un jeu de données commun. Au moment de la rédaction de ce mémoire, outre notre contribution, les équipes travaillant sur le projet sont les suivantes :

- Bureau de Recherches Géologiques et Minières (BRGM, équipes de Montpellier et Orléans).
- CEntre national du Machinisme Agricole, du Génie Rural, des eAux et des Forêts (CEMAGREF, centre d'Antony, équipe Hydrosystèmes et Bioprocédés).
- Environnements DYnamiques et TERRitoires de Montagne (EDYTEM, équipe Dynamiques Actuelles des Milieux de Montagne).
- Hydrosiences Montpellier (HSM, thème Risques Hydrologiques liés aux Aléas Extrêmes).
- Institut de Mécanique des Fluides de Toulouse (IMFT, équipe Modélisation Hydrologique Distribuée).
- Laboratoire d'étude des Transferts en Hydrologie et Environnement (LTHE).
- Météo France.

### **II.5.1 Échantillonnage**

Comme pour toute application, l'échantillonnage des signaux doit être réalisé de manière à conserver l'information, tout en diminuant la redondance que l'on aurait avec un échantillonnage trop élevé. De plus, compte tenu des inconvénients indiqués au chapitre I lorsque la complexité du réseau est trop élevée, il est nécessaire de déterminer la période d'échantillonnage permettant d'obtenir le modèle le plus parcimonieux.

Dans une approche systémique, la fréquence d'échantillonnage est fixée par la dynamique de la sortie, soit ici le débit du bassin versant. Cette fréquence doit être au moins le double de la fréquence la plus élevée contenue dans le signal (théorème d'échantillonnage de Shannon).

De manière générale, nous considérons que le processus d'accumulation des eaux de pluie au cours d'une crue est un processus intégratif. De plus, suite à des précipitations abondantes, seule une partie de la quantité d'eau tombée participe effectivement à l'écoulement engendrant la crue, le reste étant stocké dans le sol et, pour une moindre part, évapo-transpiré.

---

<sup>8</sup> Le SCHAPI dépend actuellement du Ministère chargé de l'Environnement qui est le Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer ou MEEDM.

Enfin, lorsqu'une pluie ponctuelle, assimilable à une impulsion de Dirac, se produit, elle ne donne pas lieu à des oscillations sur la sortie : le système pluie-débit peut donc être vu comme un système à réponse amortie.

Ainsi, si l'on considère que la dynamique du processus de concentration des eaux peut être approchée par celle du système linéaire intégrateur du premier ordre de constante de temps  $\tau$ , on peut exprimer le temps de réponse à 95 %, noté  $T_r$ , selon la relation bien connue :  $T_r = 3\tau$ . De plus, la sinusoïde de plus grande fréquence pouvant être transmise par ce filtre intégrateur a une période  $\tau$ .

Pour un système linéaire du premier ordre sans retard pur, le temps de réponse et le temps de montée sont proches ; pour le processus de génération des crues du Gardon, ces temps sont estimés entre 2 h et 4 h [MARCHANDISE 2007].

Il est donc possible d'encadrer le temps de réponse par ces deux valeurs :  $2h < T_r < 4h$ .

En remplaçant  $T_r$  par  $3\tau$  :  $2h < 3\tau < 4h$

D'où :  $\frac{2}{3}h < \tau < \frac{4}{3}h$ .

La réponse doit donc être échantillonnée avec une période maximale  $T_{Shannon}$  moitié, soit :  $20\text{min} < T_{Shannon} < 40\text{min}$ .

Dans la pratique une marge d'un facteur 3 à 5 est toujours prise par rapport à la fréquence théorique de Shannon, la période d'échantillonnage du débit devrait donc être inférieure à 20 minutes.

En ce qui concerne les pluies, le raisonnement précédent implique que la même période d'échantillonnage que celle des débits doit être choisie.

### **II.5.2 Cahier des charges du projet BVNE**

Le cahier des charges du projet BVNE est guidé par les caractéristiques du bassin versant et par les contraintes opérationnelles des Services de Prévision des Crues.

#### *II.5.2.a Le service Vigicrues*

La vigilance crue<sup>9</sup> est placée sous la responsabilité du MEEDADT, ce dernier devant informer le public et les personnes chargées de la gestion des crises du niveau de risque lié aux crues, estimé sur l'ensemble du réseau hydrographique sous surveillance de l'état, soit environ 20 000 km. La vigilance crue est fondée sur les mêmes principes que la vigilance météo produite par METEO FRANCE : une carte est actualisée et mise à disposition du public sur Internet (Figure 13).

Pour ce faire, chaque cours d'eau réglementaire est divisé en sections et chaque section est affectée d'une couleur indiquant le niveau de risque [VIGICRUE] :

- *Rouge : Risque de crue majeure. Menace directe et généralisée de la sécurité des personnes et des biens.*

---

<sup>9</sup> La vigilance n'est pas l'alerte : si la vigilance est identifiée par un niveau de couleur afin d'informer le public des bonnes conduites à tenir, l'alerte est déclenchée par le Préfet, si nécessaire, afin de mobiliser les moyens humains nécessaires à la gestion de crise.

- *Orange : Risque de crue génératrice de débordements importants susceptibles d'avoir un impact significatif sur la vie collective et la sécurité des biens et des personnes.*
- *Jaune : Risque de crue ou de montée rapide des eaux n'entraînant pas de dommages significatifs, mais nécessitant une vigilance particulière dans le cas d'activités saisonnières et/ou exposées.*
- *Vert : Pas de vigilance particulière*



**Figure 13 : Carte vigilance crue (exemple pour le 8 octobre 2009).**

Cette carte est accompagnée de bulletins d'information nationaux ou locaux. Ces bulletins informent sur la chronologie et le développement de l'événement de crue ; ils peuvent donner des mesures d'intensité et des prévisions pour des stations référencées. Comme pour la carte météo, des informations peuvent être obtenues localement sur un tronçon particulier (Figure 14).



Figure 14 : Carte locale des cours d'eau gérés par le SPC Grand Delta (le 8 octobre 2009)

À partir de la carte locale (Figure 14), l'utilisateur peut cliquer sur une station de mesure particulière et ainsi obtenir les mesures qui l'intéressent.

La carte *vigicrue* est mise à jour quotidiennement à 10 h et 16 h. En cas d'événement intense, elle est actualisée autant de fois que nécessaire.

#### II.5.2.b Les contraintes opérationnelles

Des contraintes fortes doivent être prises en considération du fait de l'exploitation opérationnelle des modèles de prévision. Pour qui n'a jamais vécu une crise d'inondations due à un événement cévenol, il est difficile d'imaginer que tout l'effort de travail doit se concentrer sur la prévision ; en effet, une fois l'événement cataclysmique lancé, tel celui de 2002 qui sera présenté au paragraphe II.5.3.b, l'information a beaucoup moins de valeur : tous les réseaux vont être interrompus : routiers, téléphoniques, électriques. Ainsi, si des mesures de prévention et d'organisation doivent être prises, elles doivent être diffusées *en anticipation* de l'événement ; durant l'événement ce sont malheureusement les secours qui vont demander tous les efforts.

C'est pourquoi le SCHAPI demande des prévisions jusqu'à un horizon de 5 h même si, comme nous l'avons indiqué au paragraphe II.2, le temps de concentration du bassin est plus court que cette valeur. Il faut donc délivrer des prévisions de hauteur d'eau alors que la pluie, par ailleurs difficilement prévisible, n'est pas encore tombée.

Une autre difficulté doit également être prise en considération : les incertitudes sur les données. Encore une fois, ces bassins donnant lieu à des crues rapides ont un comportement très spécifique.

Par exemple nous avons eu l'occasion d'assister à une discussion entre deux spécialistes de la prévision : le premier, habitué à des grands bassins sédimentaires revendiquait de pouvoir expertiser et valider chaque mesure appliquée à son modèle de prévision ; il faisait une

prévision à l'échelle de 24 h ; le second, travaillant sur les bassins cévenols reçoit des mesures toutes les 5 minutes pour plus d'une vingtaine de sections de cours d'eau : il n'a pas matériellement le temps de vérifier les données.

Pour cette raison, les modèles devront, à terme, avoir une certaine insensibilité aux erreurs de mesure, soit de la pluie, soit des hauteurs, car l'instrumentation peut être détériorée comme ce fut le cas en septembre 2002. On peut également envisager un fonctionnement en mode dégradé.

Par ailleurs, compte tenu des contraintes d'organisation des équipes travaillant au SPC Grand Delta, la période d'actualisation des prévisions a été définie à la  $\frac{1}{2}$  h. Nous avons vu dans le paragraphe précédent que la période définie par le théorème de Shannon était de 20 minutes ; on peut donc en conclure que l'ordre de grandeur est bon.

Enfin, comme nous l'avons signalé lors de la présentation des critères de qualité (§ II.4), le prévisionniste est plus exigeant sur la synchronisation des pics que sur l'estimation du maximum de la crue ; pour ce dernier, une erreur de 10% à 20 % est tolérée. Nous verrons au paragraphe II.5.3.e comment ces caractéristiques peuvent être dégradées au cours d'un échantillonnage insuffisant.

### **II.5.3 Base de données**

#### *II.5.3.a Courbe de tarage*

La courbe de *tarage* est la fonction non linéaire qui convertit la hauteur d'eau (variable observée) en débit. Elle est donc spécifique à un lieu particulier du cours d'eau et est sujette à des variations dans le temps du fait des changements de profil de la section du cours d'eau (creusements ou remplissage de graviers et galets, ...). La courbe de tarage est établie à l'aide de plusieurs *jaugeages* où l'on effectue simultanément une mesure de hauteur et de débit. Pour le Gardon, la courbe de tarage à Anduze a été réactualisée à partir de 2002 ; cette nouvelle courbe est reproduite en Annexe B.

Si plusieurs jaugeages ont été réalisés pour définir la courbe de tarage du Gardon à Anduze, pour des débits moyens ou faibles, en revanche, il n'existe aucun jaugeage pour les fortes valeurs et *a fortiori* pour les débits extrêmes. C'est pourquoi, afin de s'affranchir de cette source d'incertitudes, estimées supérieures ou égales à 20 ou 30% [MARCHANDISE 2007] nous choisissons la hauteur d'eau comme variable de travail de notre étude. Le réseau de neurones, effectuant une modélisation de type « boîte noire », peut tout à fait travailler sur une variable qui n'a pas le sens physique du débit.

#### *II.5.3.b Événements constitutifs*

La base de données BVNE sur laquelle nous avons travaillé comprend les données pluviométriques et limnimétriques de nombreuses stations, pour les années complètes de 1993 à 2001 échantillonnées à la période de 1 h. De cette base de données ont été extraites les hauteurs à Anduze, Saumane et Mialet ; la station de St Jean du Gard a été installée plus récemment (Figure 12). Les données pluviométriques ont été extraites pour les pluviomètres de Barre-des-Cévennes, Saint-Roman-de-Tousque, Saumane, Mialet, Soudorgues et Anduze. La base de données comprend également l'événement de septembre 2002 à l'échantillonnage de 5 min. La courbe de tarage à Anduze est fournie. En sus ont été fournies les données des

événements de 2006 et 2007, avec une période d'échantillonnage de 15 mn. Tous ces événements sont présentés graphiquement en Annexe D.

De plus il faut savoir que les mesures de l'événement de 2002 n'ont pu être réalisées en totalité car l'instrumentation a été détériorée durant l'épisode. De ce fait une étude a été commanditée au bureau d'Etude SIEE pour établir une reconstitution de cet événement à partir de laisses de crue et de témoignages [SIEE 2004]. Le bureau d'étude SIEE a donc reconstitué 2 hauteurs différentes possibles pour deux stations différentes d'Anduze, et à partir de ces hauteurs et du coefficient de rugosité du cours d'eau, a pu proposer 3 estimations de débits pour chacune des stations. Ces estimations sont également fournies dans la base de données à l'échantillonnage 5 min. Le Tableau 1 récapitule l'ensemble des événements que nous avons utilisés, et l'annexe C, leur provenance et l'échantillonnage d'origine.

N°	date	Durée (heures)	Cumul de pluies (cm)	Débit max (m <sup>3</sup> /s)	Hauteur max (m)	Débit moyen (m <sup>3</sup> /s)
1	21-24 septembre 1994	35	18,9	491	3,71	181
3	4-5 octobre 1995	54	32	975	5,34	175
4	13-14 octobre 1995	92	19	864	5	108
8	10-12 novembre 1996	82	16	268	2,71	33
13	28-29 septembre 2000	46	18	800	4,80	74
19	9 septembre 2002	29	41	2742	9,71	511
22	24-25 septembre 2006	23	6	186	2,24	21
23	19-20 octobre 2006	55	20	1436	6,61	181
24	17 novembre 2006	34	6	275	2,75	65
25	20-23 novembre 2007	70	12	264	2,69	139
107	5-7 novembre 1997	74	27	624	4,20	225
108	26-27 novembre 1997	66	11	244	2,58	117
109	18-19 décembre 1997	104	38	985	5,37	327
117	20-21 octobre 1999	34	12	473	3,64	212
120	12-14 novembre 2000	71	13	279	2,77	110

**Tableau 1 : Liste des événements et leurs caractéristiques hydrologiques**

À cause de son amplitude exceptionnelle, l'événement de 2002 est choisi par le SCHAPI comme l'événement prototype de test des modèles de prévision. Pour cette raison, tous les résultats de prévision seront présentés sur cet événement.

Il apparaît sur le Tableau 1 que l'événement le plus intense de la base est l'événement 19 de 2002 ; avec un débit de pointe estimé à 2742 m<sup>3</sup>/s, cet événement a été la cause de 22 décès et de 1,2 milliard d'euros de dégâts économiques [LE LAY *et al.* 2007]. L'annexe D illustre l'ampleur de cette catastrophe régionale, et il faut remonter jusqu'à l'année 1958 pour trouver un événement d'une telle amplitude. Le second événement par ordre d'amplitude est l'événement 23 de 2006, mais il atteint un débit deux fois moindre. On peut également

remarquer que l'événement 19 est parmi les plus courts de la base : il fait donc réellement partie des événements rapides.

### *II.5.3.c Typologies*

Comme nous l'avons présenté dans le chapitre précédent sur l'apprentissage des réseaux de neurones, la base des exemples disponibles doit être séparée en plusieurs sous-ensembles, utilisés soit pour l'apprentissage soit pour le test du modèle. Au sein de l'ensemble d'apprentissage, on peut distinguer également plusieurs sous-ensembles qui seront autant d'ensembles utilisés pour la validation croisée. Enfin, si l'on souhaite utiliser la méthode de régularisation par arrêt précoce, il faut également définir un ensemble d'arrêt.

Le comportement du bassin versant au cours d'un événement pluvieux intense étant dynamique, il faut conserver la structure temporelle de l'événement ; pour cette raison, les sous-ensembles ne peuvent être constitués que d'événements entiers. Or seuls 15 événements sont disponibles : il est donc exclu de les affecter aléatoirement dans les sous-ensembles. Il faudra au contraire choisir cette répartition de manière logique. Ainsi, pour répartir au mieux les événements dans les sous-ensembles d'apprentissage, de validation, d'arrêt et de test, nous avons procédé à une recherche de typologie des événements pluvieux.

Après plusieurs tentatives fondées sur des critères hydrologiques ou statistiques, nous avons choisi la typologie suivante : trois critères de discrimination suivants ont été définis :

- Événements intenses.
- Événements longs.
- Événements complexes (plusieurs pics).

À partir de ces trois critères, 8 classes peuvent être définies, qui sont repérées par une lettre par ordre alphabétique de A à F (Tableau 2). Il apparaît en particulier que la catégorie de l'événement 19, du 9 septembre 2002, ne comporte que deux éléments : le 19 et le 13. Certaines catégories sont vides d'éléments.

Un événement est considéré comme intense si son pic atteint la moitié du pic de l'événement le plus intense (ici l'événement 19) ; de même, un événement est considéré comme long si sa durée dépasse la moitié de l'événement le plus long, et il est qualifié de multi-pics s'il comporte plusieurs pics bien dissociables.



N°	Intense	Long	Mono-pic	Classe
19	oui	non	oui	C
23	oui	oui	oui	A
109	oui	oui	oui	A
3	oui	oui	non	B
4	oui	oui	oui	A
13	oui	non	oui	C
107	non	oui	non	F
1	non	non	non	H
117	non	non	oui	G
120	non	oui	non	F
24	non	non	oui	G
8	non	oui	non	F
25	non	oui	non	F
108	non	oui	non	F
22	non	non	oui	G

**Tableau 2 : Classification des événements**

À partir de cette classification, il aurait été idéal de disposer d'un représentant de chaque classe dans chaque sous-ensemble d'apprentissage, de test, et d'arrêt. À la lecture du Tableau 3, qui présente le nombre d'événements appartenant à chaque classe, il apparaît clairement que ceci n'est pas possible ; ainsi, il n'y a qu'un représentant de la classe B, c'est-à-dire des événements intenses, longs et avec plusieurs pics (événement d'octobre 1995, qui pourrait d'ailleurs être séparé en deux événements distincts).

Intense	Long	Mono-pic	Classe et nombre d'éléments
oui	oui	oui	A (3)
oui	oui	non	B (1)
oui	non	oui	C (2)
oui	non	non	D (0)
non	oui	oui	E (0)
non	oui	non	F (5)
non	non	oui	G (3)
non	non	non	H (1)

**Tableau 3 : Répartition des événements dans les classes**

Il apparaît donc clairement que la base de données des événements, même si elle comporte une quantité d'information importante et précieuse, va poser de sérieuses difficultés pour la conception d'un système à apprentissage statistique.

#### *II.5.3.d Instrumentation du bassin versant du Gardon d'Anduze*

Les pluies et débits sont mesurés à l'aide de sept pluviomètres répartis sur le bassin versant (Barre-des-Cévennes, Saint-Roman-de-Tousque, Saumane, Mialet, Soudorgues, Saint-Jean-du-Gard et Anduze), et de trois stations de mesure de hauteur à Mialet, Saumane et Anduze (Figure 12). Les chroniques de Mialet et Saumane présentant de nombreuses lacunes, c'est la station d'Anduze qui a été choisie pour effectuer la prévision des hauteurs d'eau ; la station de Saint-Jean-du-Gard n'est pas utilisée car elle n'est en fonction que depuis 2000.

#### *II.5.3.e Bruits de mesures*

Les signaux mesurés lors des crues éclair sont malheureusement entachés d'une grande imprécision. Celle-ci a plusieurs origines. La plus importante est sans doute l'hétérogénéité des précipitations : en effet, la mesure des précipitations est réalisée au moyen de pluviomètres, qui, s'ils font partie des dispositifs de mesure des précipitations les plus précis, ont néanmoins une précision quantifiée à 10% ou 20% [MARCHANDISE 2007]. Les pluviomètres échantillonnent le cumul de pluie et le télétransmettent toutes les 5 minutes, mais la maille du réseau est insuffisante pour garantir que de fortes précipitations ne soient pas tombées entre 2 pluviomètres, et demeurent ainsi non mesurées. Par exemple, lors de l'événement de 2002, on a pu relever un cumul près de trois fois plus important à Soudorgues qu'à Anduze, alors que les deux stations ne sont distantes que d'une quinzaine de kilomètres (Figure 12). Pour remédier à cet inconvénient, les dernières recherches proposent d'utiliser les estimations des précipitations issues des radars avec une définition d'un km<sup>2</sup>, mais cette technologie ne peut pas actuellement fournir des cartes, pour les événements disponibles, avec un traitement homogène de la réflectivité<sup>10</sup>. Cette solution ne peut donc être choisie pour cette étude : nous supposons donc que l'échantillonnage spatial à l'aide de 6 pluviomètres répartis sur le bassin versant est suffisant [OBLED, 2008]. Au niveau des sorties, la hauteur d'eau peut être mesurée, mais l'estimation du débit par la courbe de tarage est en revanche beaucoup moins maîtrisée, du seul fait de l'absence de jaugeage durant les épisodes de pluies intenses ; il est en effet matériellement dangereux de s'approcher du cours d'eau en de telles circonstances ! Comme nous l'avons déjà indiqué, nous nous affranchissons de cette imprécision en travaillant sur la variable hauteur d'eau.

La fréquence d'échantillonnage des signaux de sortie est également génératrice d'erreurs ; en effet, les événements mesurés de 1994 à 2002 sont échantillonnés toutes les heures, contre 5 minutes à partir de 2002. Or il apparaît, sur l'événement de septembre 2002, que l'échantillonnage à 1 h fait perdre une quantité non négligeable d'information, comme indiqué dans le Tableau 4. Il faut noter que, si l'échantillonnage est réalisé avec une fréquence insuffisante, l'écart entre l'instant relatif au début de la période d'échantillonnage et celui du pic de la courbe a une forte influence sur l'intensité du pic et sur sa position.

---

<sup>10</sup> Un projet de ré-analyse des lames d'eau RADAR est actuellement mené par METEO FRANCE afin de constituer une archive de données de précipitations avec un échantillonnage d'une heure et sur un maillage de 1 km<sup>2</sup> à partir de 1997.

Non seulement l'amplitude du pic peut être atténuée de 17%, mais la prévision de l'heure du pic peut être entachée d'un retard d'une heure. Une analyse de la variance a montré par ailleurs que la période maximale d'échantillonnage devrait être de 15 à 30 minutes. Enfin, rappelons que les hauteurs d'eau de l'événement de septembre 2002 (le plus intense de la base de données), sont le résultat d'une estimation réalisée *a posteriori*.

Période d'échantillonnage	5 mn	15 mn	½ h	1 h	2 h
Débit de pointe (m3/s)	3030	2869	2742	2508	2358
Erreur induite		5%	9%	17%	22%
Période du maximum de débit	6h	6h	6h30	7h	7h
Retard induit par l'échantillonnage	0	0	30min	1h	1h

**Tableau 4 : Influence de l'échantillonnage sur l'amplitude et le déphasage du pic principal de l'événement de septembre 2002.**

Ainsi tous les événements antérieurs à 2002 sont entachés d'une erreur due à l'échantillonnage, erreur qui dépend de la dynamique propre de l'événement et que nous ne pouvons quantifier *a posteriori*.

Il est clair que, comme la dimension réduite de la base d'apprentissage, toutes ces erreurs d'estimation ou de mesure contribuent à rendre la tâche plus difficile au système devant réaliser la prévision des crues.

## **II.6 Conclusion**

Après avoir présenté les Gardonnades, et leurs caractéristiques de crues éclair, puis précisé les enjeux de la modélisation de tels événements, nous nous sommes intéressés à la modélisation hydrologique de ces phénomènes. Les modèles étudiés permettent tous de considérer que la répartition spatiale des précipitations est la variable principale qui permet d'appréhender le processus de genèse de ces crues. L'examen de la base de données nous a permis de mettre en évidence la disponibilité actuelle d'un nombre suffisant de pluviomètres échantillonnés avec une fréquence convenable, même si l'échantillonnage temporel des événements ayant eu lieu avant 2002 est insuffisant. Cette difficulté, ajoutée au nombre d'événements réduits de la base de données, ainsi qu'aux erreurs et imprécisions de mesures importantes, nous a permis de prendre la mesure du défi qui est posé aux systèmes à apprentissage statistique pour parvenir à prévoir, en l'absence de prévision de pluie, l'évolution de l'événement de septembre 2002, deux fois plus important que le deuxième événement le plus intense, en terme d'intensité du pic de crue, de la base de données (*i.e.* l'événement de 2006). Le chapitre suivant présente la méthodologie qui a permis de relever ce défi en utilisant un prédicteur à réseaux de neurones.

## **II.7 Références bibliographiques du Chapitre II**

- [AYRAL 2005] P-A. AYRAL. "Contribution à la spatialisation du modèle opérationnel de prévision des crues éclair ALHTAÏR : Approches spatiale & expérimentale – Application au bassin versant du Gardon d'Anduze". Thèse de l'Université Aix-Marseille I, France, Spécialité Biosciences de l'Environnement, 300p., Mars 2005
- [BESSIERE *et al.* 2007] H. BESSIERE, H. ROUX, D. DARTUS. « Data assimilation and distributed flash flood modeling ». In: Second Space for Hydrology Workshop, Surface Water Storage and Runoff: Modeling, In-situ data and Remote Sensing, 12 - 14 Novembre 2007, Genève, Suisse.
- [BOUTTIER 2003] F. BOUTTIER. "The Arome mesoscale project". Proceedings of the ECMWF September 2003
- [BORRELL 2004] V. ESTUPINA BORRELL. "Vers une modélisation hydrologique adaptée à la prévision opérationnelle des crues éclair - Application à de petits bassins versants du sud de la France". Thèse, Institut National Polytechnique de Toulouse. E. D. Sciences de l'Univers, de l'Environnement et de l'Espace, Spécialité Hydrologie, 228 p., Avril 2004
- [BOUVIER *et al.* 2006] C. BOUVIER, P.-A. AYRAL, P. BRUNET, A. MARCHANDISE, C. MARTIN. « Recent advance in rainfall-runoff modelling : extrapolation to extreme floods un southern France » International Workshop on Hydrological Extermes. Rende 3-4, May 2006
- [HEWLETT 1961] J. D. HEWLETT « Watershed management » Annu. Rep., U.S. Dep. of Agric for Serv., Southeast., For Exp. Stn., Asheville, N.C., Journal of Hydrology, pp. 61-66, 1961
- [HORTON 1933] R. E. HORTON. "The role of infiltration in the hydrological cycle". Trans. Am. Geophys. Union, vol 14, pp. 446-460, 1933
- [KITADINIS 1980] P. K. KITADINIS, R. BRAS. "Real-Time forecasting with a conceptual hydrologic model, applications and results". Water Resources Research, vol 16, n° 6, pp. 1034-1044, 1980
- [LE LAY 2006] M. LE LAY. « Modélisation hydrologique dans un contexte de variabilité hydro-climatique. Une approche comparative pour l'étude du cycle hydrologique à Meso-échelle du Bénin ». Thèse de l'INPG, 2006
- [LE LAY *et al.* 2007] M. LE LAY, G. M. SAULNIER. « Exploring the signature of climate and landscape spatial variabilities in flash flood events : Case of the 8-9 September 2002 Cévennes-Vivarais catastrophic event ». Geophysical Research Letters, vol. 34, L13401 5 pages, 2007
- [MARCHANDISE 2007] A. MARCHANDISE. "Modélisation hydrologique distribuée sur le Gardon d'Anduze ; étude comparative de différents modèles pluie-débit, extrapolation de la normale à l'extrême et tests d'hypothèses sur les processus hydrologiques". Université Montpellier II, Sciences et Techniques du Languedoc, Thèse soutenue le 26 mars 2007
- [MICHEL 1983] C. MICHEL. "Que peut-on faire en hydrologie avec un modèle à réservoir ?". La Houille blanche, n° 1, pp. 39-44, 1983

- [NASH *et al.* 1970] J. E. Nash, J. V. SUTCLIFFE. "River Flow Forecasting through Conceptual Models. Part I – A Discussion of Principles". Journal of Hydrology, vol 10, pp. 282-290, 1970
- [OBLED *et al.* 2008] C. OBLED, I. ZIN, B. HINGRAY. « Choix des pas de temps et d'espace pour des modélisations parcimonieuses en hydrologie des crues ». Colloque SHF - *Prévisions hydrométéorologiques* – pp. 127-134, 2008
- [SAULNIER *et al.* 1997] G-M. SAULNIER, K. BEVEN, C. OBLED. « Including spatially variable effective soil depths in TOPMODEL ». Journal of Hydrology, vol 202, pp. 158-172, 1997
- [SAULNIER *et al.* 2009] G-M. SAULNIER, M. LE LAY. « Sensitivity of flash °flood simulations on the volume, the intensity and the localization of rainfall in the Cévennes-Vivarais region (France) ». Water Resources Research. *A paraître* 2009
- [SIEE 2004] SIEE – Société d'Ingénierie Eau et Environnement. « Validation des relevés hydrométriques de l'événement des 8 & 9 septembre 2002 ». Rapport d'Etude, Direction Départementale de l'Équipement du Gard, 162 p., 2004
- [VIGICRUE] <http://www.vigicrues.ecologie.gouv.fr/>. Extrait de la carte le 8 octobre 2009.

# **III Application des réseaux de neurones à la prévision des crues éclair du bassin versant d'Anduze**

## **III.1 Introduction**

Le premier chapitre de ce mémoire nous a permis de présenter non seulement les propriétés de la modélisation par apprentissage statistique à l'aide de réseaux de neurones, mais également ses limitations, notamment la nécessité de disposer d'une base d'apprentissage suffisamment vaste et représentative du comportement du processus à modéliser. En effet, la possibilité d'identifier la fonction sous-jacente au processus, indépendamment de la réalisation particulière du bruit présent dans les exemples utilisés en apprentissage, nécessite une base d'exemples riche et complète.

Le chapitre II a exposé le problème des crues éclair des Gardons et a insisté sur le peu d'événements disponibles ainsi que sur les sources de bruits et d'imprécisions opérant tant en entrée qu'en sortie du processus. Ce dernier chapitre nous a permis de présenter le défi qui est posé par la modélisation des Gardonnades en l'absence de prévision de pluie ; c'est ce défi qui est relevé dans le présent chapitre.

Ainsi, nous présenterons en premier lieu un aperçu des travaux réalisés sur la prévision des crues, notamment des crues éclair, avec les réseaux de neurones. Nous pourrions constater que faute d'avoir appréhendé le problème de la généralisation avec toutes ses implications, nombre des études concluent sur l'insuffisance, ou l'inadéquation, de la modélisation statistique en hydrologie.

C'est pourquoi la deuxième partie de ce chapitre explique comment nous avons appliqué des méthodes de régularisation, parfois simultanément, pour remédier tant au petit nombre d'événements, qu'à la présence de bruit important. Ainsi ont été évalués la validation croisée, l'arrêt précoce et la modération des poids. Si la validation croisée et la modération des poids s'appliquent assez directement, en revanche le choix de l'ensemble d'arrêt nécessaire à l'arrêt précoce s'est révélé délicat.

Finalement, une méthodologie a été établie, qui a permis d'effectuer des prévisions en aveugle d'excellente qualité, à notre connaissance inégalée, sur l'événement majeur de septembre 2002. Cette méthodologie prouve que lorsque le problème de la généralisation est bien posé, et que des méthodes rigoureuses sont appliquées, l'apprentissage statistique est tout à fait à même de réaliser la synthèse d'un prédicteur de crue opérationnel en temps réel.

## **III.2 Mise en œuvre des réseaux de neurones pour la prévision**

### **III.2.1 Prévision des crues par réseaux de neurones**

Les travaux sur l'utilisation des réseaux de neurones pour effectuer la prévision des crues ont commencé au début des années 1990 [DARTUS *et al.* 1993], [JOHANNET *et al.* 1994], [KARUNANITHI *et al.* 1994]. En effet la relation pluie-débit, et généralement les phénomènes hydrologiques sont considérés comme dynamiques, non linéaires et souvent difficiles à modéliser ; l'application de l'apprentissage statistique à ce champ scientifique est

donc logique. L'engouement suscité par la thématique *apprentissage statistique appliqué à l'hydrologie* a ainsi donné lieu à plusieurs articles de synthèse [COULIBALY *et al.* 1999], [GOVINDARAJU 2000 a&b] et à de nombreuses études, telles celle de [ZEALAND *et al.* 1999] pour la prévision des débits à l'échéance de la semaine sur un grand bassin versant de l'Ontario (Canada), ou la comparaison des performances des prédicteurs de processus statiques ou dynamiques<sup>11</sup> [COULIBALY *et al.* 2000]. De manière générale, la principale propriété mise en avant est la propriété d'approximation universelle : c'est pourquoi le plus grand nombre des travaux s'intéresse au perceptron multicouche et utilise des méthodes d'optimisation de la fonction de coût telles que l'algorithme de Levenberg-Marquard ou la descente de gradient simple. En revanche, très peu de travaux examinent rigoureusement la capacité de généralisation, et comment les méthodes de régularisation peuvent aider à éviter le surajustement, même si de nombreuses études butent sur le problème de la généralisation. Ainsi [PIOTROWSKY *et al.* 2006] étudie la prévision des crues rapides sans aucune méthode de régularisation dans le cadre de plusieurs modèles non linéaires autorégressifs et note que « *tous les perceptrons multicouches étudiés ont des performances significativement plus mauvaises lorsqu'ils sont appliqués à l'ensemble de test* », ce qui était parfaitement prévisible et aurait pu être évité. [COULIBALY *et al.* 2000 b] utilise l'arrêt précoce, et [SUDHEER *et al.* 2002] la validation croisée pour sélectionner la complexité, mais aucun des deux n'évalue la généralisation sur un ensemble indépendant des données utilisées en apprentissage, en validation, ou en arrêt ; la qualité de leurs résultats est donc vraisemblablement sur-évaluée. [GAUME *et al.* 2003] s'interroge sur l'effet du surparamétrage des réseaux de neurones par rapport aux modèles hydrauliques conceptuels qui, comme indiquée au chapitre II, possèdent peu de paramètres. L'analyse présentée par E. Gaume *et al.* est intéressante car elle pose le problème de la spécialisation que réalise l'apprentissage par rapport au bruit présent dans les données d'apprentissage (phénomène fondamental bien connu sous le nom de *dilemme biais-variance*), mais, contrairement aux règles méthodologiques bien établies, les auteurs n'utilisent pas de méthode de régularisation pour résoudre ce problème ; c'est ce qui explique leur conclusion : « *Les modèle de type boîtes noires, tels les réseaux de neurones, ne peuvent distinguer entre le bruit et la variance* ».

Plus spécifiquement en ce qui concerne les crues éclair, très peu de travaux existent sur l'utilisation de l'apprentissage statistique. Nous avons déjà cité [PIOTROWSKY *et al.* 2006] qui n'effectue pas de régularisation. [SCHMITZ *et al.* 2008], qui effectue la prévision des crues de la rivière Freiberger Mulde, affluente de l'Elbe (Est de l'Allemagne), considère que les réseaux de neurones « *... sont prisonniers de leurs données d'apprentissage* » et contourne le problème de la mauvaise généralisation en plaçant dans l'ensemble d'apprentissage l'événement le plus intense de sa base de données. [SAHOO *et al.* 2006 a&b] effectue la prévision de crues éclair à Hawaï ; il n'applique pas de méthodes de régularisation, mais obtient de bons résultats. En effet, comme la prévision est réalisée pour de très petits bassins versants (2 ou 15 km<sup>2</sup>), il y a vraisemblablement peu d'erreur d'estimation dans les

---

<sup>11</sup> Une confusion est souvent réalisée entre le processus étudié et le prédicteur associé. Ainsi cet article ne traite que de prédicteurs statiques, mais qui identifient un processus dynamique. Le prédicteur choisi est du type *dirigé*, qui comporte en entrée, outre les entrées exogènes, la mesure de la sortie du processus. Le prédicteur utilisé par Coulibaly n'est donc pas récurrent.

mesures de précipitation, du fait du peu d'hétérogénéité spatiale, et le réseau de neurones s'ajuste correctement à la fonction pluie-débit, en l'absence de bruit.

Enfin, du point de vue hydrologique, seul le bassin versant de la rivière Freiberger Mulde, étudiée par SCHMITZ *et al.* est comparable au bassin versant du Gardon par ses pentes moyennes et ses dimensions ; néanmoins, les précipitations y sont moins intenses, l'intensité du phénomène est donc moindre même si les dégâts peuvent y être également catastrophiques. On peut noter que SCHMITZ *et al.* travaillent également en l'absence de prévision de pluie car les phénomènes météorologiques opérant en montagne sont difficiles à prévoir. Dans les deux cas le temps de réponse du bassin et la fréquence d'échantillonnage des signaux sont du même ordre de grandeur (horaire).

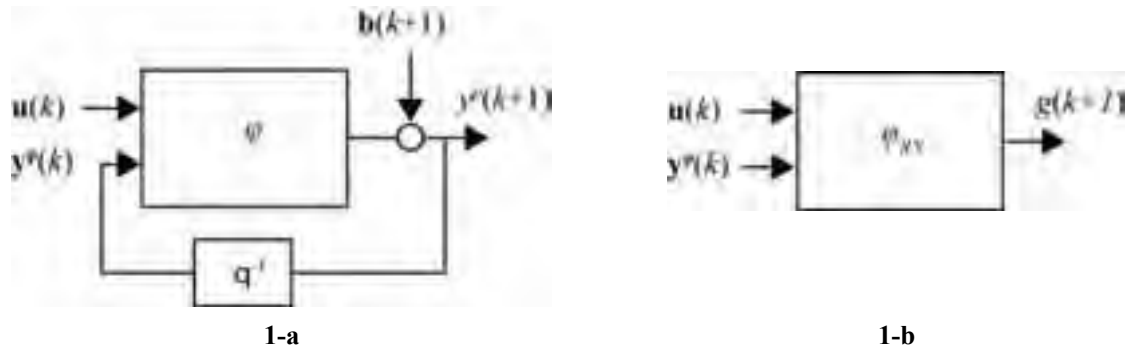
En résumé, il apparaît que le problème de l'aptitude des prédicteurs neuronaux à généraliser est largement évoquée dans la littérature sur la prévision des crues, mais que les réponses apportées (qui sont parfois la recommandation d'abandonner les techniques statistiques) ne sont pas toujours adéquates. Pour la prévision des crues éclair, à notre connaissance, aucune méthodologie n'a donné de résultats efficaces sur des phénomènes aussi intenses que ceux des bassins versants cévenols. L'objet de ce chapitre est donc de montrer comment l'application d'une procédure rigoureuse de conception de modèle, comprenant l'optimisation de la complexité, la sélection des variables, l'intégration de méthodes de régularisation durant l'apprentissage, et une évaluation sur un ensemble de test réellement indépendant de la conception du modèle, conduit à des prévisions satisfaisantes d'événements extrêmes.

### **III.2.2 Modèle de prédicteur proposé**

La relation pluie-débit étant dynamique et non linéaire, le choix du prédicteur associé est réalisé comme indiqué au chapitre I, en section I.4. Le prédicteur choisi doit être associé à un modèle de processus postulé, en considérant le mode d'action des perturbations. Dans cette étude, les deux types de représentation entrée-sortie, avec les hypothèses bruit d'état et bruit de sortie ont été testés, en évaluant les performances de leurs prédicteurs associés : le prédicteur dirigé et non dirigé.

De manière assez prévisible, le prédicteur dirigé, correspondant à l'hypothèse *bruit d'état* est apparu comme le plus performant car il estime la sortie du processus à partir des valeurs précédentes de ce dernier : cette tâche est plus simple à réaliser que celle du prédicteur non-dirigé qui estime la sortie du processus à partir des estimations, entachées d'erreur, de sa propre sortie. Nous allons donc nous attacher à présenter la démarche et les résultats obtenus avec le prédicteur dirigé. Les représentations du modèle postulé et du prédicteur associé, présentées en Figure 9 du chapitre I, sont reprises ci-dessous afin de faciliter la compréhension du lecteur.





**Figure 15 : Modèle postulé (1-a) de la relation pluie-débit, et prédicteur neuronal dirigé (1-b) associé à l'hypothèse *bruit d'état*.**

Le vecteur des entrées exogènes  $\mathbf{u}(k)$ , à l'instant présent  $k$ , est constitué des précipitations observées sur les 6 pluviomètres installés sur le bassin versant (chapitre II.2), sur une fenêtre temporelle  $f_T$  qui permet donc de prendre en considération les précipitations avec un historique de profondeur  $f_T$ .

Le vecteur des sorties observées du processus  $\mathbf{y}^p(k)$  est également appliqué en entrée du prédicteur avec une profondeur temporelle notée  $r$ .

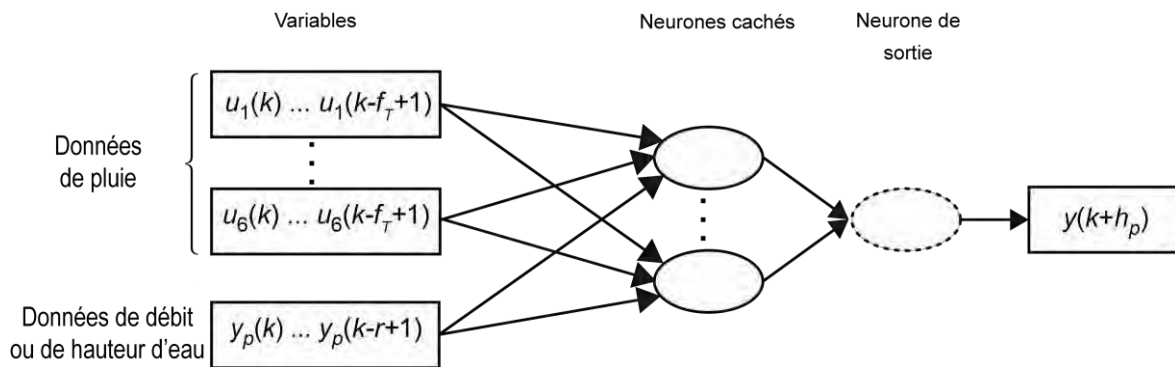
La relation réalisée par le prédicteur est donc la suivante :

$$g(k+h_p) = \varphi_{RN}(\mathbf{y}^p(k), \dots, \mathbf{y}^p(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-f_T+1), \mathbf{w})$$

où :

- $h_p$  est l'horizon de prévision du prédicteur,
- $y^p(k+h_p)$  est la grandeur que l'on veut prédire, mesurée à l'instant  $k+h_p$ ,
- $\mathbf{u}(k)$  est le vecteur des variables exogènes à l'instant  $k$ .

Le réseau de neurones choisi pour réaliser la fonction  $g(k)$  étant un perceptron multicouches, le prédicteur avec ses entrées peut donc être représenté comme sur la Figure 16, où  $y(k+h_p)$  est la sortie estimée par le prédicteur.



**Figure 16 : Le type de prédicteur choisi.**

Les entrées exogènes  $u_i(k)$  sont appliquées avec un historique de profondeur  $f_T$  à définir, ainsi que les valeurs observées du processus avec un historique de profondeur  $r$ .

### III.3 Sélection des variables

#### III.3.1 Dimensionnement de la fenêtre temporelle de pluie $f_T$

La variable exogène principale du prédicteur est la variable de précipitation. Nous avons souligné, dans le chapitre II, l'importance de l'information spatialisée de pluie, et justifié le fait que celle-ci pouvait être prise en considération au moyen de 6 pluviomètres répartis sur le bassin versant. Comme la hauteur (et donc le débit) à un instant  $k$  ne dépend pas de la pluie qui tombe à cet instant mais de celle qui est tombée aux instants précédents (compte tenu de la durée de l'acheminement de l'eau), il convient de déterminer la profondeur nécessaire de l'historique des pluies. Cette profondeur est représentée par la fenêtre temporelle de pluie  $f_T$ . L'ajustement de  $f_T$ , comme pour les autres variables, sera réalisé au moyen de méthodes de sélection de variables.

#### III.3.2 Variables supplémentaires

##### III.3.2.a Cumul de pluie

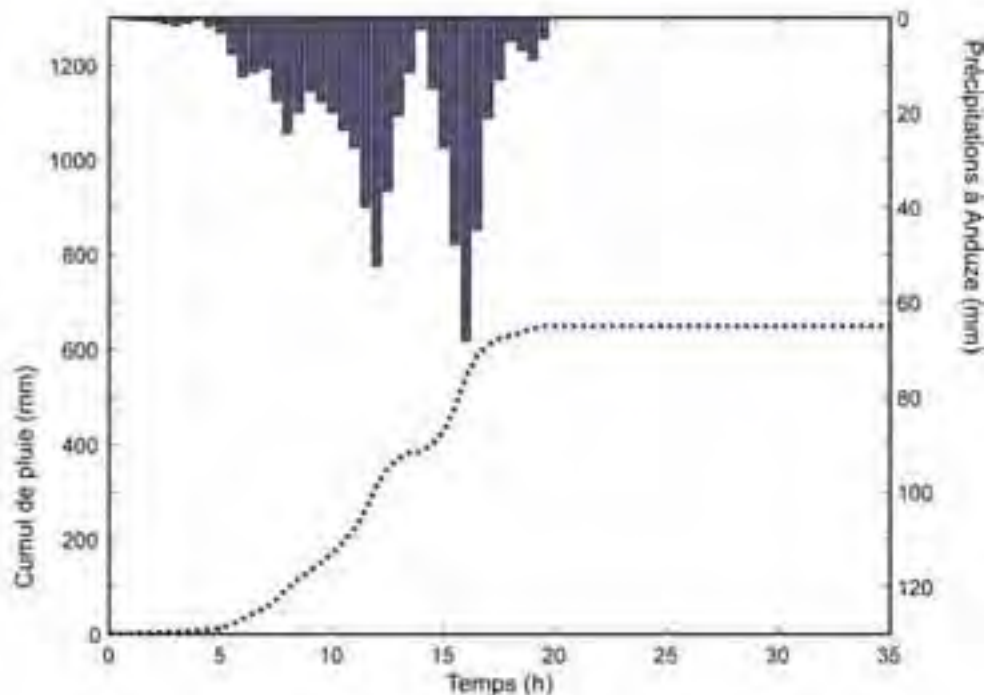
Nous avons rappelé, dans le chapitre I, que le surajustement du réseau de neurones est favorisé par une trop grande complexité, c'est-à-dire un nombre excessif de paramètres ; c'est pourquoi la dimension de l'historique des précipitations sera choisie aussi faible que possible. Néanmoins, il pourrait également être utile d'ajouter une information sur le cumul des précipitations, qui transcrirait alors de manière très compacte l'historique observé depuis le début de l'événement. Grâce à ce cumul, les variables comportent alors une information à long terme.

Le cumul est calculé ainsi :

$$\Pi(k) = \sum_{l=0}^k p(l)$$

où  $p(l)$  est le cumul des précipitations tombées durant l'intervalle de temps  $l$ .

Au niveau hydrologique, un cumul de pluie élevé au moment du pic des précipitations peut laisser présager une saturation du sol importante ; *a contrario*, un cumul de pluie faible impliquerait que le sol n'est pas saturé. À titre d'illustration, la Figure 17 présente l'évolution du cumul du pluviomètre d'Anduze pour l'événement de septembre 2002. On remarque que deux fronts pluvieux se sont succédés durant cet événement : le premier a un cumul de 382 mm, avec un pic de 52 mm (sur ½ heure), et le second apporte un autre cumul de 265 mm avec un pic de 68 mm (sur ½ heure). On peut imaginer que, lors de l'arrivée du second front pluvieux, l'information de cumul de précipitations issues du front précédent peut être utile. On peut également noter qu'un cumul élevé n'est pas nécessairement lié à une intensité de pic importante ; en particulier, dans les deux exemples précédents l'intensité du premier front pluvieux, plus faible que celle du second front, conduit néanmoins à un cumul plus important pour le premier front pluvieux que pour le second.



**Figure 17 : Evolution du cumul des précipitations de l'événement de septembre 2002 à Anduze**

Le cumul des précipitations est tracé en ordonnées positives (échelle de gauche) tandis que les précipitations sont en ordonnées inversées (échelle de droite), échantillonnées à la demi-heure.

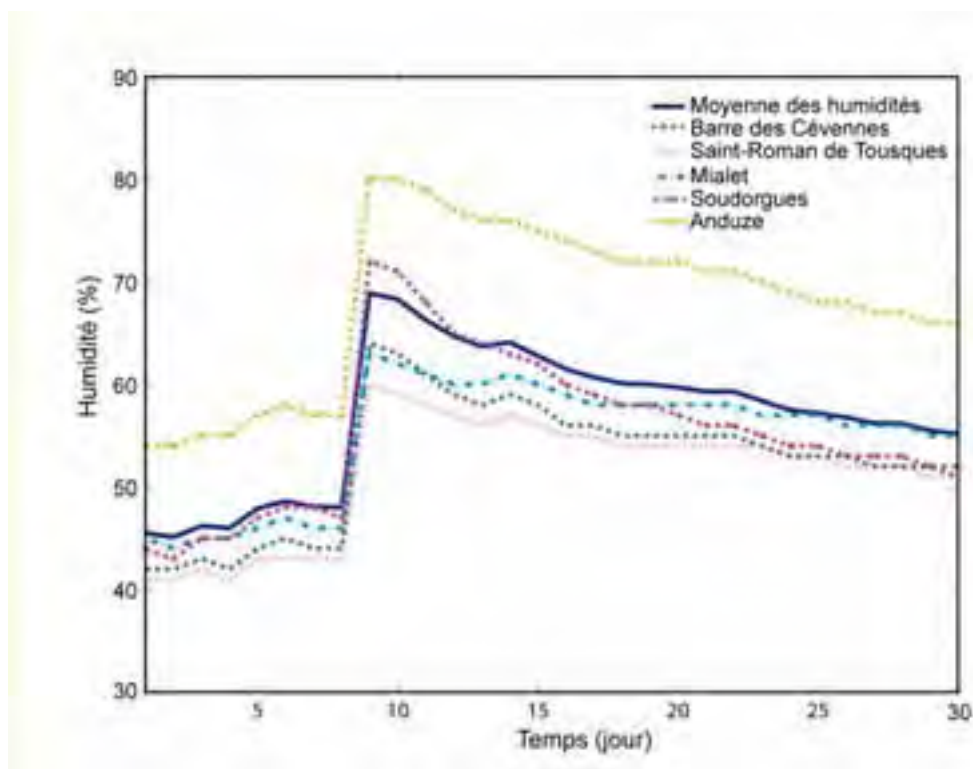
Les méthodes de sélection de variable qui seront décrites dans la section III.5 ont néanmoins démontré que la variable cumul n'était pas nécessaire au bon fonctionnement du prédicteur.

### *III.3.2.b Humidité des sols*

L'humidité des sols représente l'état hydrique du sol ; elle affecte directement l'évaporation réelle, la répartition de la pluie entre infiltration et ruissellement, et la percolation vers les couches de sol profondes. Cette information est importante dans les modèles hydrologiques, et nous avons vu au chapitre II qu'elle est estimée comme essentielle pour le modèle TOPMODEL : son importance est classée juste après celle de la spatialisation des précipitations.

L'humidité des sols dont nous disposons pour cette étude (base de données BVNE) provient du modèle ISBA, «Interaction Sol Biosphère Atmosphère ». ISBA a été développé dans le but de simuler les échanges d'eau et d'énergie entre le sol, la végétation et les basses couches de l'atmosphère. Pour établir son bilan hydrique, ISBA calcule plusieurs grandeurs telles que la quantité d'eau interceptée par la végétation, l'équivalent en eau du manteau neigeux total, la quantité d'eau présente dans le sol, ainsi que les flux d'eau ruisselée, drainée, évaporée, par la surface du sol nu d'une part, et par la végétation d'autre part. L'humidité des sols fournie par ISBA est une information journalière. Sa dynamique est peu, voire non adaptée à la prévision des crues éclair qui présentent des temps de montée en crue bien plus rapides, de l'ordre de 2 heures. Ainsi, l'évolution de l'humidité des sols estimée par ISBA est représentée sur la Figure 18 pour le mois de septembre 2002 : on peut noter que, du fait de son actualisation

journalière, cette humidité augmente brusquement à minuit, sans relation avec l'humidité réelle des sols.



**Figure 18 : Humidité ISBA lors de l'événement de septembre 2002**

On peut noter sur la Figure que l'humidité augmente brutalement entre les jours 8 et 9, c'est-à-dire entre le 8 et le 9 septembre 2002. Ce saut n'est évidemment pas conforme à l'évolution réelle de l'humidité dans les sols.

La méthode que nous présenterons dans la section III.6.1.2 a mis en évidence le fait que cette variable ne permet pas d'améliorer la prévision des crues. Deux explications sont proposées à ce résultat :

- d'une part la mauvaise adéquation de l'échantillonnage ISBA avec la dynamique des crues, que nous venons d'illustrer en commentaire de la figure précédente,
- d'autre part, lorsque le prédicteur est du type dirigé comme représenté sur la Figure 16, les évolutions mesurées de la hauteur d'eau tiennent déjà compte de l'humidité réelle du sol : le réseau a donc déjà l'information sous forme implicite dans les variables.

Cette variable ne sera donc pas conservée pour effectuer la prévision.

### **III.4 Sélection de modèle par validation croisée**

#### **III.4.1 Application de la validation croisée**

##### *III.4.1.a Rappel de la méthode*

La validation croisée, présentée dans la section I.3.7 permet d'estimer l'erreur de généralisation du modèle ; elle opère à partir de l'ensemble de validation, et il est possible de

sélectionner la complexité du réseau à l'aide du score de validation croisée. Rappelons que ce dernier est défini par :

$$S = \sqrt{\frac{1}{D} \sum_{E_p=1}^D J(E_p)^2},$$

$$\text{avec } J(E_p) = \frac{1}{N_p} \sum_{m=1}^{N_p} (y^m - y(\mathbf{x}^m, \mathbf{w}))^2,$$

où :

- $E_p$  est le sous-ensemble d'exemples numéro  $p$  ( $p = 1$  à  $D$ ) comportant  $N_p$  exemples<sup>12</sup>,
- $J_p$  est la fonction de coût calculée sur le sous-ensemble  $E_p$ , en général l'erreur quadratique moyenne minimale sur l'ensemble de validation en faisant varier l'initialisation des paramètres du réseau,
- $y^m$  et  $y(\mathbf{x}^m, \mathbf{w})$  sont les valeurs mesurée et estimée de la grandeur à prédire, pour le sous-ensemble de validation  $m$ .

Il est également possible de calculer le score de validation croisée en prenant en considération d'autres critères que l'erreur quadratique, en particulier à partir du coefficient de détermination et du critère de persistance.

Pour le critère de persistance, on définit le score de validation croisée par :

$$S_p = \frac{1}{D} \left[ \sum_{E_p=1}^D C_p(E_p) \right],$$

où  $C_p$  est le critère de persistance calculé sur le sous-ensemble de validation  $E_p$ .

Pour le critère de Nash :

$$S_d = \frac{1}{D} \left[ \sum_{E_p=1}^D R^2(E_p) \right],$$

où  $R^2(E_p)$  est le coefficient de détermination calculé sur le sous-ensemble de validation  $E_p$ .

Dans le chapitre I, nous avons présenté la validation croisée comme méthode de régularisation passive utilisée pour sélectionner la complexité du réseau. À cette fin, elle sera utilisée pour effectuer la sélection des variables et ajuster le nombre de neurones cachés du prédicteur. Sont ainsi déterminés par validation croisée :

- la profondeur  $f_T$  de l'historique de pluies,
- le nombre  $r$  d'observations précédentes de la sortie du processus,
- le nombre  $N_c$  de neurones cachés,

---

<sup>12</sup> En général lorsque l'on applique la validation croisée, les ensembles de validation sont définis de tailles identiques. Dans cette application comme les sous-ensembles de validations sont des événements pluvieux ils ont des tailles différentes, il est donc nécessaire de réduire chaque erreur par la taille de l'échantillon observé.

- l'initialisation des paramètres,
- l'opportunité de l'ajout des variables cumul et humidité.

Pour alléger la lecture de ce manuscrit, lorsque la sélection de l'ensemble de ces grandeurs sera effectuée, nous parlerons de la sélection du modèle ; lorsque la sélection d'une seule de ces grandeurs sera envisagée, nous préciserons quelle sélection est en cours (par exemple sélection de la complexité pour le choix du nombre de neurones cachés).

Une fois la validation croisée effectuée et le modèle choisi, l'apprentissage est réalisé avec tous les événements sur lesquels s'est déroulée la validation croisée ; l'estimation de la généralisation est réalisée ensuite.

Notons que la validation croisée est appliquée sur la base de données comportant 13 événements sur les 15 disponibles ; en effet, les crues de septembre 2000 (événement 13) et de septembre 2002 (événement 19) ne sont jamais présentées en apprentissage ou en validation, afin de pouvoir établir des comparaisons homogènes en fonction des différentes méthodes proposées. L'événement de septembre 2002 sera systématiquement utilisé pour évaluer la généralisation sur un ensemble indépendant de l'ensemble d'apprentissage ; l'événement de septembre 2000 sera utilisé par la suite comme ensemble d'arrêt.

L'apprentissage est effectué par calcul du gradient par l'algorithme de rétropropagation, puis optimisation de la fonction de coût, à l'aide du gradient ainsi calculé, par l'algorithme de Levenberg-Marquardt dont le principe a été présenté au chapitre I, et qui est exposé en Annexe A. Compte tenu de la taille de la base de données disponible, cet algorithme est généralement considéré comme le plus efficace des algorithmes d'optimisation connus à ce jour.

Sans autre méthode de régularisation, la fin de chacune des phases d'apprentissage est déterminée par un nombre maximal de présentations soit 50 itérations.

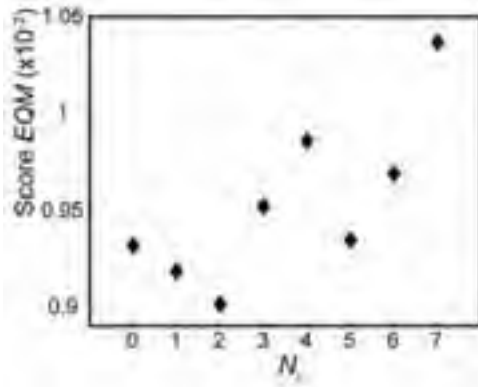
Nous effectuons 100 initialisations des paramètres en raison de la grande variabilité observée en fonction de cette dernière.

#### *III.4.1.b Exemple : sélection de la complexité*

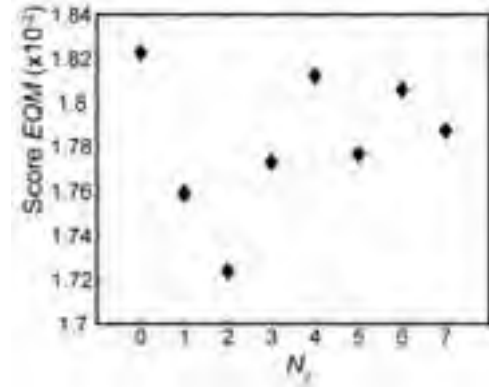
Le score minimum (pour toutes les initialisations) de l'erreur quadratique moyenne est reporté en fonction du nombre de neurones cachés  $N_c$ . De manière classique, nous avons procédé par augmentation progressive de la complexité du modèle. Le modèle le plus simple est le modèle linéaire, il est constitué d'un seul neurone linéaire ; il est reporté sur la Figure 19 comme le modèle à 0 neurone caché. Ensuite le nombre de neurones cachés, variant de 1 à 7 est reporté dans les graphes de la Figure 19.

Nous remarquons sur la Figure 19 que, quel que soit l'horizon de prévision, l'introduction du premier neurone caché améliore le score de l'erreur quadratique moyenne par rapport à celui du réseau linéaire ; ceci signifie que la relation que nous cherchons à identifier est bien non linéaire.

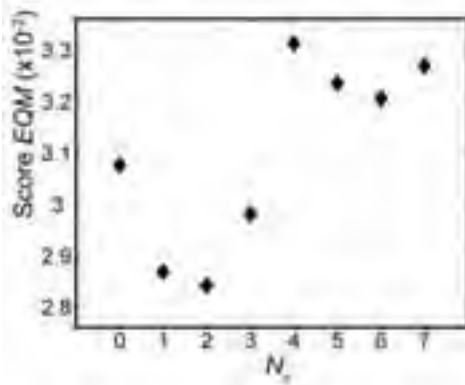
Au-delà de 4 neurones cachés, les scores se dégradent nettement ; nous avons choisi 7 neurones cachés comme complexité maximale.



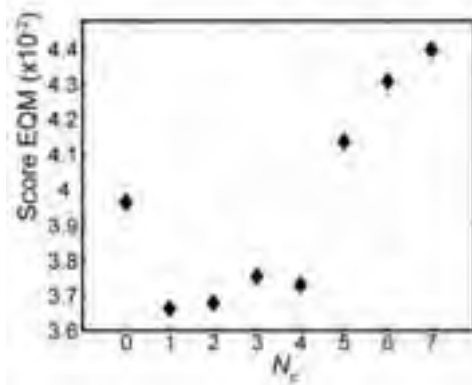
$h_p=0,5$  h



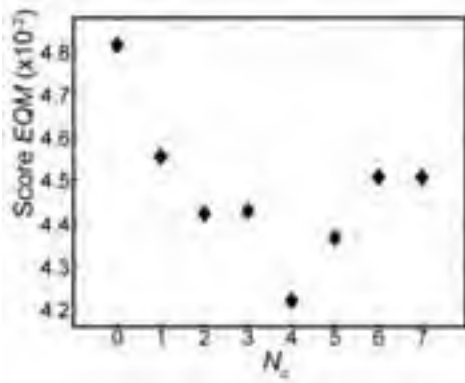
$h_p=1$  h



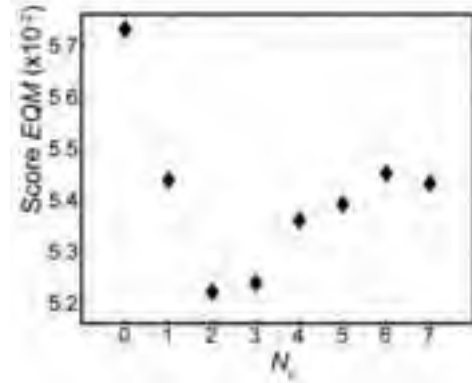
$h_p=2$  h



$h_p=3$  h



$h_p=4$  h



$h_p=5$  h

**Figure 19 : Evolution du score minimum de l'erreur quadratique moyenne pour chaque horizon de prévision.**

#### *III.4.1.c Meilleurs prédicteurs obtenus*

Le Tableau 5 présente, de manière synthétique, les résultats obtenus.

L'horizon de prévision  $h_p$ , exprimé en heures, représente l'intervalle de temps, à partir de l'instant courant  $k$ , au-delà duquel on cherche une estimation de la sortie du processus.

Pour chacun des critères et suivant les 100 initialisations différentes, nous avons calculé : le score minimum, le score moyen, l'écart-type du score et le score médian.

Horizon de prévision $h_p$ (heures)		0,5	1	2	3	4	5
Fenêtre temporelle $f_T$ (heures)		2,5	3	3	2	0,5	0,5
Nombre de neurones cachés $N_c$		2	2	2	1	4	2
Retard $r$ sur $y^p$		2	2	2	2	2	2
Erreur Quadratique Moyenne EQM	Score minimum ( $\times 10^{-2}$ )	0,9	1,72	2,84	3,66	4,22	5,22
	Score moyen ( $\times 10^{-2}$ )	1,61	2,24	4,62	3,97	4,98	5,66
	ET (score) ( $\times 10^{-2}$ )	4,03	1,07	8,52	0,84	0,87	0,22
	Score médian ( $\times 10^{-2}$ )	1,08	2,01	3,22	3,68	4,84	5,64
Critère de hauteur $C_h$	Score maximum	0,95	0,93	0,88	0,85	0,82	0,78
	Score moyen	0,94	0,90	0,85	0,84	0,80	0,76
	ET (score) ( $\times 10^{-2}$ )	0,9	1	1	2,42	1,24	0,86
	Score médian	0,94	0,91	0,85	0,85	0,80	0,76

**Tableau 5 : Meilleurs modèles issus de la validation croisée seule**

En regardant les valeurs des écarts-types, on peut tout d'abord noter sur le Tableau 5 une grande variabilité des valeurs de l'erreur quadratique, et de hauteur en fonction de l'initialisation. L'écart type peut même être supérieur à la valeur moyenne, ce qui s'explique par une distribution des valeurs très asymétrique, en l'occurrence avec quelques valeurs très élevées. Pour obtenir une bonne erreur quadratique, il est donc primordial d'initialiser les paramètres du réseau convenablement.

Le Tableau 5 permet d'obtenir les meilleures configurations du prédicteur pour chaque horizon de prévision ; on pourra remarquer que l'historique des précipitations diminue lorsque l'horizon de prévision augmente ; on pourrait même y relever une relation approximative du type  $f_T=5,5-h_p$ , qui signifierait que l'information des précipitations a une influence sur la hauteur d'eau pendant une durée maximale de 5,5 heures.

Le nombre de neurones cachés est égal en général à 2, mais varie entre 1 et 4. Lorsque ce dernier est égal à 1, on peut se poser la question de la linéarité de la relation pluie-débit prédite ; nous verrons dans la suite que le réseau multicouche est bien nécessaire, car le réseau possédant un seul neurone linéaire a également été évalué, et se révèle insuffisant.

Le nombre d'entrées de hauteur mesurée est toujours égal à 2 : ceci signifie que le prédicteur a besoin des mesures de hauteurs aux deux instants précédents l'instant actuel  $k$ . A partir de



celles-ci, le prédicteur peut déduire la vitesse de variation de hauteur afin, par exemple d'en extraire l'humidité du sol.

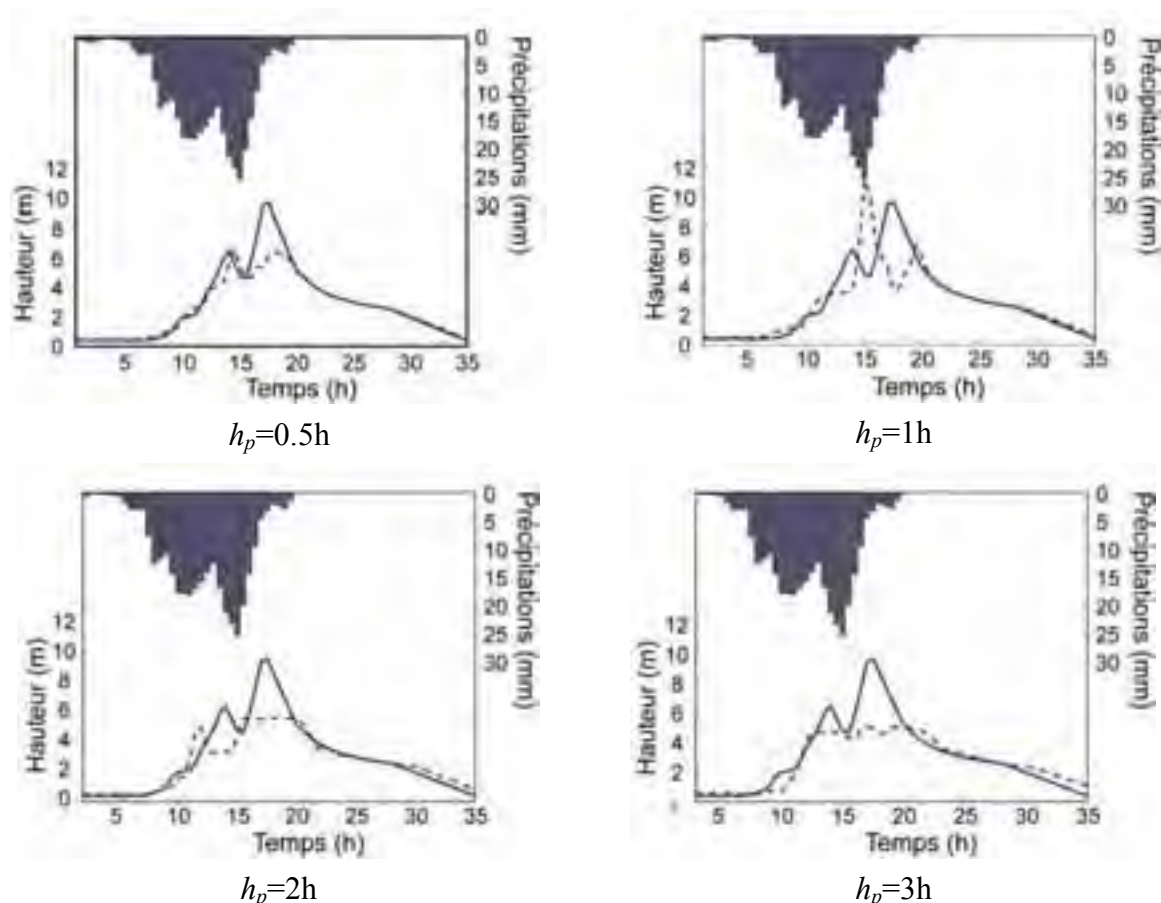
#### *III.4.1.d Test sur la crue de septembre 2002*

Une fois les meilleurs modèles sélectionnés et l'apprentissage réalisé, il est nécessaire d'évaluer la généralisation sur un événement indépendant. Comme nous l'avons indiqué précédemment, l'événement 19 de septembre 2002 a été mis de côté à cette fin.

Compte tenu des éléments de bibliographie présentés au début de ce chapitre, il est particulièrement ambitieux de tester les performances du prédicteur sur l'événement de septembre 2002, qui est l'événement le plus intense de la base de données. En effet sa hauteur maximale (9,6 m) est de 50% supérieure à celle du second événement le plus intense : celui d'octobre 2006 (6,6 m). Néanmoins, compte tenu des pertes humaines et des dégâts matériels occasionnés par l'événement de 2002, il nous a paru que le véritable enjeu de cette étude est bien de prévoir cet événement majeur ; c'est pourquoi nous l'avons choisi comme événement de test.

Pour chaque horizon de prévision, les limnigrammes prévus sont reportés sur la Figure 20 pour chacune des meilleures configurations (indiquées dans le Tableau 5).

Par une analyse visuelle, nous pouvons noter que ces prévisions sont peu satisfaisantes. Les pics de crue prévus ne suivent pas du tout les observations. À un horizon d'une demi-heure, la prévision sous-estime de près de 40% le pic de la crue, tandis qu'à une heure d'horizon de prévision, la prévision n'est plus synchronisée avec l'observation.



**Figure 20 : Limnigrammes de test avec la validation croisée seule**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Aux horizons de prévision lointains, c'est-à-dire à partir de deux heures et plus, on note une saturation autour de 6 m de hauteur. Or 6 m de hauteur d'eau est approximativement la hauteur atteinte par le second événement le plus intense, celui d'octobre 2006. Cet événement est présenté au réseau lors de l'apprentissage ; ainsi, nous pouvons supposer que, sans une méthode de régularisation supplémentaire, le réseau de neurones ne serait pas capable de généraliser au-delà des valeurs qu'il a rencontrées en phase d'apprentissage.

De manière plus synthétique, les valeurs des critères de qualité pour les différents horizons de prévision sont reportées dans le Tableau 6.

Les valeurs négatives du critère de persistance présentes dans le Tableau ci-dessous signifient que les limnigrammes prévus sont plus mauvais que la prévision naïve (cette dernière suppose que la hauteur ne change pas entre la valeur actuelle et la valeur prédite). Ainsi, en moyenne, la prévision naïve serait plus fiable que la prévision par le prédicteur ainsi réalisé !

En revanche, les coefficients de détermination  $R^2$  (critère de Nash) aux différents horizons de prévision semblent corrects, même pour les 2 premiers horizons de prévision, à savoir 0,5 h et 1 h. Cette observation, apparemment contradictoire avec l'analyse précédente des coefficients de persistance, illustre bien le défaut du critère de Nash que nous avons souligné lors de sa

présentation : comme ce critère est fondé sur l'erreur quadratique moyenne, une translation horizontale faible de la courbe prévue<sup>13</sup>, ou quelques erreurs localisées, peuvent ne pas influencer beaucoup sa valeur. On obtient alors des valeurs satisfaisantes, alors que l'analyse visuelle de la prévision est très médiocre ; on peut noter que l'inverse est également possible.

Les critères de hauteur viennent confirmer la constatation visuelle faite sur les limnigrammes prévus. Sur les différents horizons de prévision, on peut noter que les prévisions atteignent en moyenne seulement 63% des pics observés.

En conclusion, ces résultats, obtenus sans application de méthodes de régularisation, sont médiocres et nous nous proposons de les améliorer en appliquant des méthodes de régularisation actives présentées dans le chapitre I.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	2	2	1	4	2	2
Critères de Persistance $C_p$	-1,74	-1,85	0,47	0,60	0,37	0,49	-0,28
$R^2$ (Critère de Nash)	0,88	0,55	0,77	0,74	0,42	0,32	0,61
Critère de hauteur $C_h$	0,74	0,65	0,65	0,69	0,54	0,49	0,63

Tableau 6: Critères de qualité de la prévision avec la validation croisée seule

### III.5 Impact des méthodes de régularisation actives

#### III.5.1 Régularisation par arrêt précoce

##### III.5.1.a Définition de l'ensemble d'arrêt

Lorsque l'arrêt précoce est utilisé (section I.3.6.b), il faut définir un ensemble d'arrêt sur lequel l'erreur quadratique est estimée durant la phase d'apprentissage ; lorsque cette erreur atteint son minimum, l'apprentissage est arrêté afin d'éviter le surajustement.

Lorsque la base des exemples est très étendue, on choisit habituellement un ensemble d'exemples de manière aléatoire afin de constituer un ensemble d'arrêt qui représente statistiquement le même comportement que l'ensemble complet de la base des exemples. Or, pour la prévision des crues du Gardon, comme nous l'avons souligné au chapitre II, nous ne

<sup>13</sup> A titre d'illustration, comme indiqué au chapitre II, nous proposons en Annexe F le calcul des coefficients de détermination de la prévision naïve ; cette prévision, qui ne présente aucun intérêt pour le prévisionniste, conduit néanmoins à des coefficients de détermination élevés !

disposons que de 15 événements de crues ; l'affectation aléatoire de quelques événements dans un ensemble d'arrêt n'est donc pas envisageable. C'est pour cette raison que nous avons défini la typologie présentée au chapitre II (section II.5.3.c) ; elle propose de considérer 3 caractéristiques permettant de définir 8 classes : les événements intenses, longs, ou possédant plusieurs pics.

Ainsi, parmi toutes celles qui étaient possibles, trois façons de constituer l'ensemble d'arrêt nous ont paru logiques et ont été étudiées :

- ensemble d'arrêt « complexe ». L'ensemble d'arrêt, pour représenter l'ensemble des comportements possibles, est constitué à partir des classes : il doit comporter un représentant de chacune des classes. Il apparaît immédiatement que ceci n'est pas envisageable car il y a 8 classes et seulement 15 événements ; on ne peut pas utiliser la moitié des événements comme ensemble d'arrêt. A défaut, l'ensemble d'arrêt est constitué avec un représentant de chacune des caractéristiques (intense, long, multi-pic).
- l'ensemble d'arrêt et l'ensemble de validation sont identiques. L'ensemble d'arrêt ne comporte qu'un événement, mais tous sont utilisés successivement afin de définir le meilleur modèle moyen. Ceci est fait en effectuant l'arrêt sur l'ensemble utilisé pour la validation, lorsque la validation croisée est utilisée.
- ensemble d'arrêt « simple ». L'ensemble d'arrêt ne comporte qu'un événement, toujours le même, représentatif du comportement moyen de la base des exemples.

#### *III.5.1.b Ensemble d'arrêt « complexe »*

Lorsque l'ensemble d'arrêt est constitué à partir des classes de manière à prendre en considération un comportement « complexe », nous affectons à l'ensemble d'arrêt trois crues aux caractéristiques différentes : une crue intense, une crue à plusieurs pics et une crue de longue durée.

De nombreuses combinaisons d'événements ont été essayées mais, comme nous allons l'illustrer sur l'exemple suivant, il apparaît que la fonction demandée au réseau est trop complexe. En effet, pour trouver un compromis entre les trois types de crues, l'apprentissage dure longtemps, et l'arrêt n'est déclenché qu'après 30 ou 50 présentations de l'ensemble d'apprentissage. Ceci est contradictoire avec l'objet de la méthode ; arrêter l'apprentissage assez tôt pour éviter le surajustement.

Sur la Figure 21, nous avons représenté la prévision effectuée sur un ensemble d'arrêt particulier. Ici l'ensemble d'arrêt est constitué des 3 événements suivants :

- événement 4 d'octobre 1995 : intense (0-38 h de la Figure 21),
- événement 8 de novembre 1996 : multi-pic (39-120 h de la Figure 21),
- événement 120 du 12-14 novembre 2000 : long (121-188 h de la Figure 21).

L'apprentissage est alors réalisé sur tous les autres événements de la base, hormis l'événement de septembre 2002. Rigoureusement, on ne peut pas mesurer les performances de généralisation sur un ensemble tel que l'ensemble d'arrêt puisque ce dernier est utilisé pour

arrêter l'apprentissage ; il est néanmoins intéressant d'observer le comportement du prédicteur sur cet ensemble<sup>14</sup>.

Ainsi, sur la Figure 21, il apparaît que, pour un horizon de prévision de 2 heures, les limnigrammes prévus sont excellents : le coefficient de détermination est très élevé (0,93), et le critère de hauteur atteint 0,85 de moyenne sur tous les pics. Le critère de persistance est positif (0,64) ; de plus, les pics observés et simulés sont synchrones, et les erreurs d'anticipation se produisent surtout lors de la décrue, ce qui est moins gênant pour les prévisionnistes.

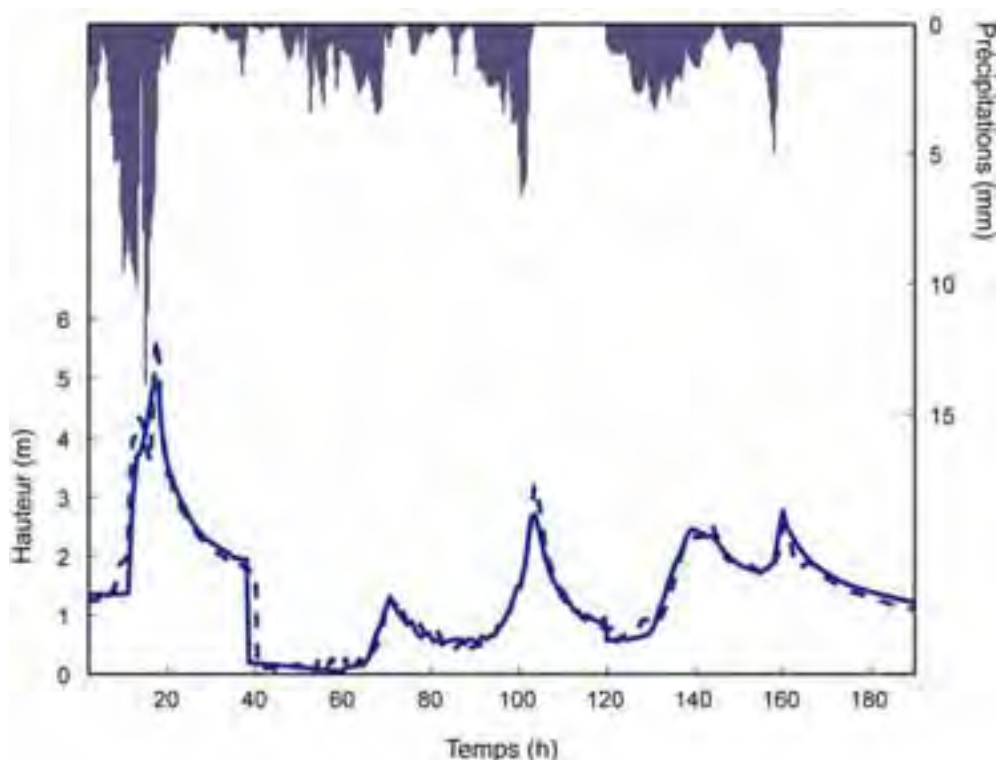


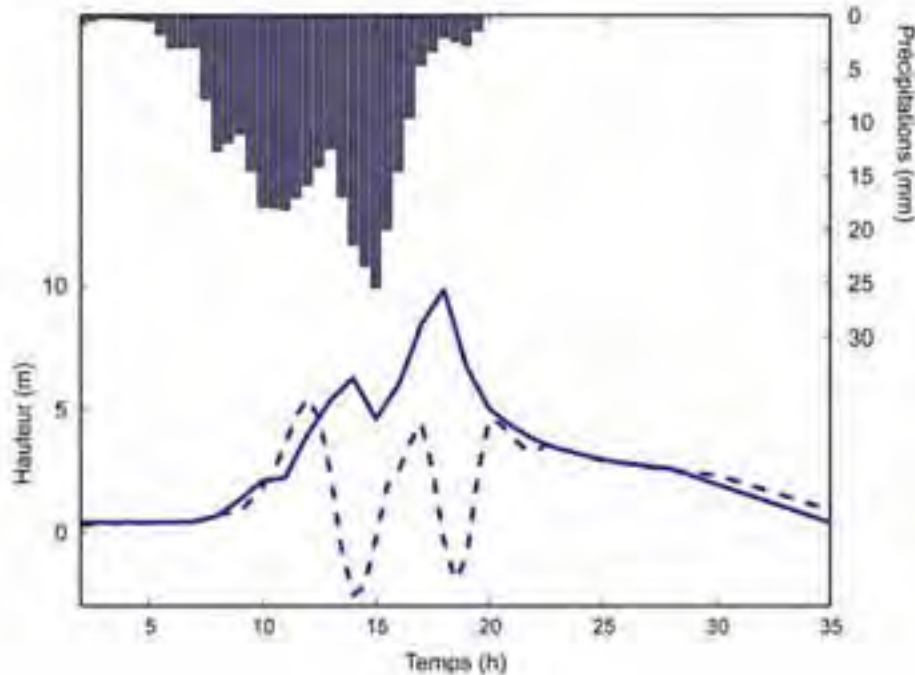
Figure 21 : Prédiction sur l'ensemble d'arrêt constitué de 3 événements ( $h_p=2h$ )

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Dans cette configuration, le test sur l'événement de 2002 conduit à des résultats extrêmement médiocres ( $R^2 = -0,57$ ,  $C_p = -2,76$  et  $C_h$  quasiment nul) : le réseau est incapable de généraliser correctement (Figure 22). Le choix d'un ensemble d'arrêt complexe conduit visiblement à un surajustement du réseau.

De manière évidente, les performances observées sur l'ensemble d'arrêt sont bonnes alors que celles observées sur l'ensemble de test sont mauvaises : on ne peut donc évaluer correctement la généralisation sur l'ensemble d'arrêt ; **le réseau effectue une spécialisation très forte par rapport à l'ensemble d'arrêt.**

<sup>14</sup> Lorsque l'arrêt précoce est utilisé, l'estimation de la généralisation sur l'ensemble d'arrêt est pourtant ce qui est réalisé dans la majorité des publications en hydrologie.



**Figure 22 : Généralisation à l'horizon 2 h pour l'événement de septembre 2002**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur. On peut noter que la hauteur prévue atteint même des valeurs négatives.

#### *III.5.1.c L'ensemble d'arrêt et l'ensemble de validation sont identiques*

Le paragraphe précédent a permis de mettre en évidence le fait que l'ensemble d'arrêt devait être assez « simple » pour que le réseau réalise l'apprentissage rapidement afin de mettre en œuvre réellement un arrêt précoce permettant d'éviter le surajustement. À cette fin, nous proposons d'utiliser un ensemble d'arrêt comportant un seul événement. Cependant pour éviter la spécialisation par rapport à ce seul événement, mise en évidence dans la section précédente, nous proposons que l'ensemble de validation soit utilisé tant pour la validation croisée que pour l'ensemble d'arrêt.

La procédure est la suivante :

- dans un premier temps, la validation croisée est appliquée de manière à sélectionner le modèle ; en parallèle, l'arrêt de l'apprentissage est fait par l'arrêt précoce sur l'ensemble utilisé pour la validation. Le nombre moyen de présentations effectuées en apprentissage avant l'arrêt est calculé (7 présentations dans ce cas),
- une fois le modèle sélectionné, l'apprentissage est à nouveau réalisé, cette fois avec tous les événements sauf ceux de septembre 2000 et septembre 2002, l'arrêt est réalisé à l'itération moyenne d'arrêt observée lors de la phase précédente et le test peut être réalisé sur l'événement de septembre 2002.

Tout d'abord, comme l'arrêt moyen de l'apprentissage est réalisé après seulement 7 présentations, l'objectif de réaliser un arrêt précoce est bien obtenu. Nous présenterons les résultats complets de cette procédure à la section III.5.1.e.

*III.5.1.d L'ensemble d'arrêt ne comporte qu'un événement, représentatif du comportement moyen de la base des exemples*

L'ensemble d'arrêt est l'événement de 2002

En préambule à cette section, et pour établir des limnigrammes de référence, nous allons choisir comme ensemble d'arrêt « simple » l'événement de septembre 2002. Compte tenu de la spécialisation observée par l'ensemble d'arrêt, le prédicteur sera spécialisé sur la crue de septembre 2002. Si le test est également effectué sur l'événement de 2002, nous pourrions ainsi disposer de limnigrammes de référence que l'on peut penser être les meilleurs possibles.

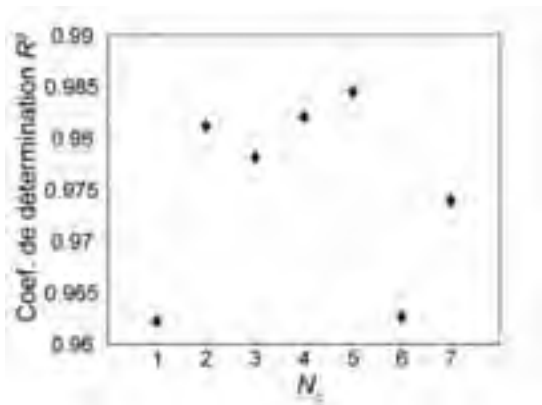
La procédure est simplifiée : pour chaque horizon de prévision, 100 initialisations sont réalisées. Pour chacune on effectue l'arrêt précoce sur l'événement de 2002 et le meilleur test sur ce même événement détermine le prédicteur choisi. On a donc, dans ce cas, les mêmes ensembles pour l'arrêt et le test, en l'absence de validation croisée. Cette procédure n'est pas statistiquement correcte, mais rappelons qu'il ne s'agit pas de réaliser un prédicteur utilisable sur le terrain, mais d'établir des limnigrammes de référence.

Tout d'abord, la complexité doit être déterminée ; pour cela, nous présentons sur la Figure 23 l'évolution du coefficient de détermination en fonction du nombre de neurones cachés pour chaque horizon de prévision. Par exemple pour l'horizon 2 h, cette évolution indique que 4 neurones cachés doivent être choisis, le  $R^2$  est alors de 0,93. Il correspond à un  $R^2$  excellent, comme on peut le constater sur le limnigramme présenté en Figure 24. Toujours pour 2 h, le critère de persistance est excellent (0,83), le pic principal est très bien prévu, et celui de hauteur est excellent (0,90).

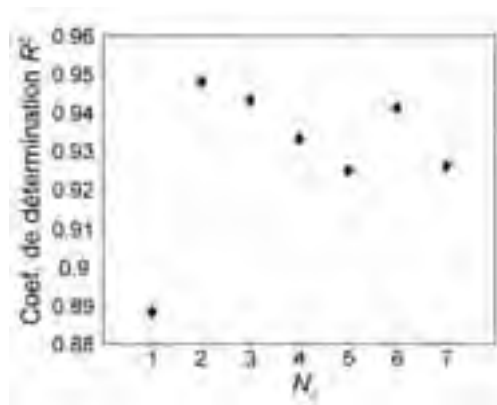
Une fois la complexité déterminée, le test est réalisé sur l'événement de 2002, les limnigrammes obtenus sont présentés sur la Figure 24.

Les limnigrammes présentés sur la Figure 24 sont, comme nous l'avons signalé, les meilleurs limnigrammes possibles avec la complexité et les variables choisies. Il est donc intéressant de noter que dès l'horizon de prévision 4h, les résultats se détériorent et deviennent médiocres pour 5 h. On peut donc dès à présent penser que, en l'absence de prévision de pluie, on ne pourra obtenir de bonnes prévisions pour l'horizon de prévision 5 h : le prédicteur n'a pas encore l'information de pluie, car celle-ci n'est pas tombée, quand on lui demande de prévoir le débit 5 h à l'avance.

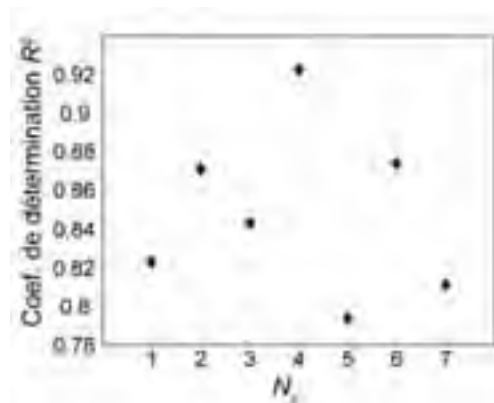
Au niveau des critères de qualité, on peut noter sur le Tableau 7 que les trois critères de qualité sont bons, même pour l'horizon de prévision de ½ heure, pour lequel les pics estimé et observé sont synchrones.



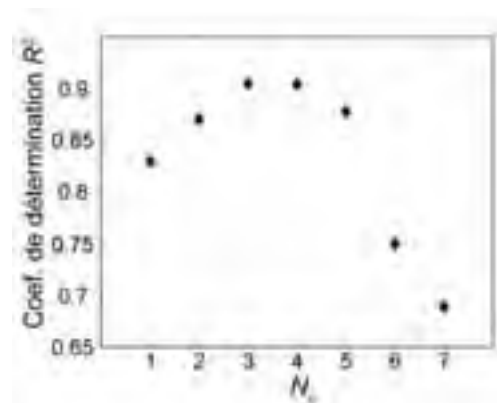
$h_p=0,5$  h



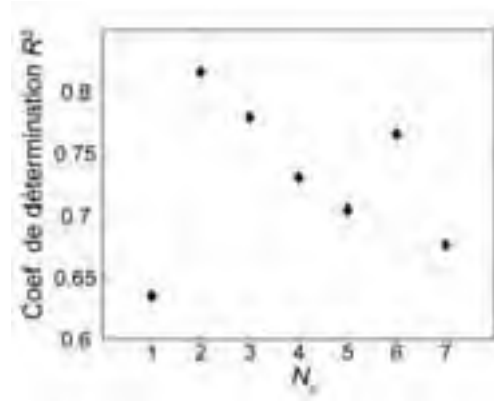
$h_p=1$  h



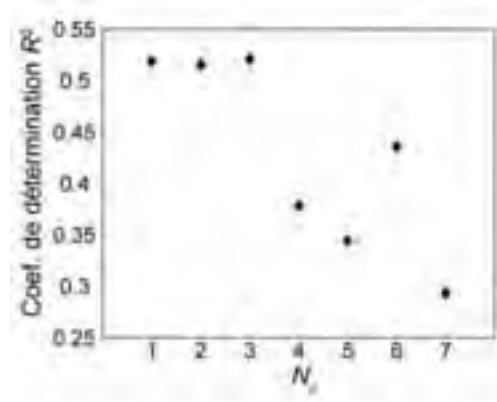
$h_p=2$  h



$h_p=3$  h



$h_p=4$  h

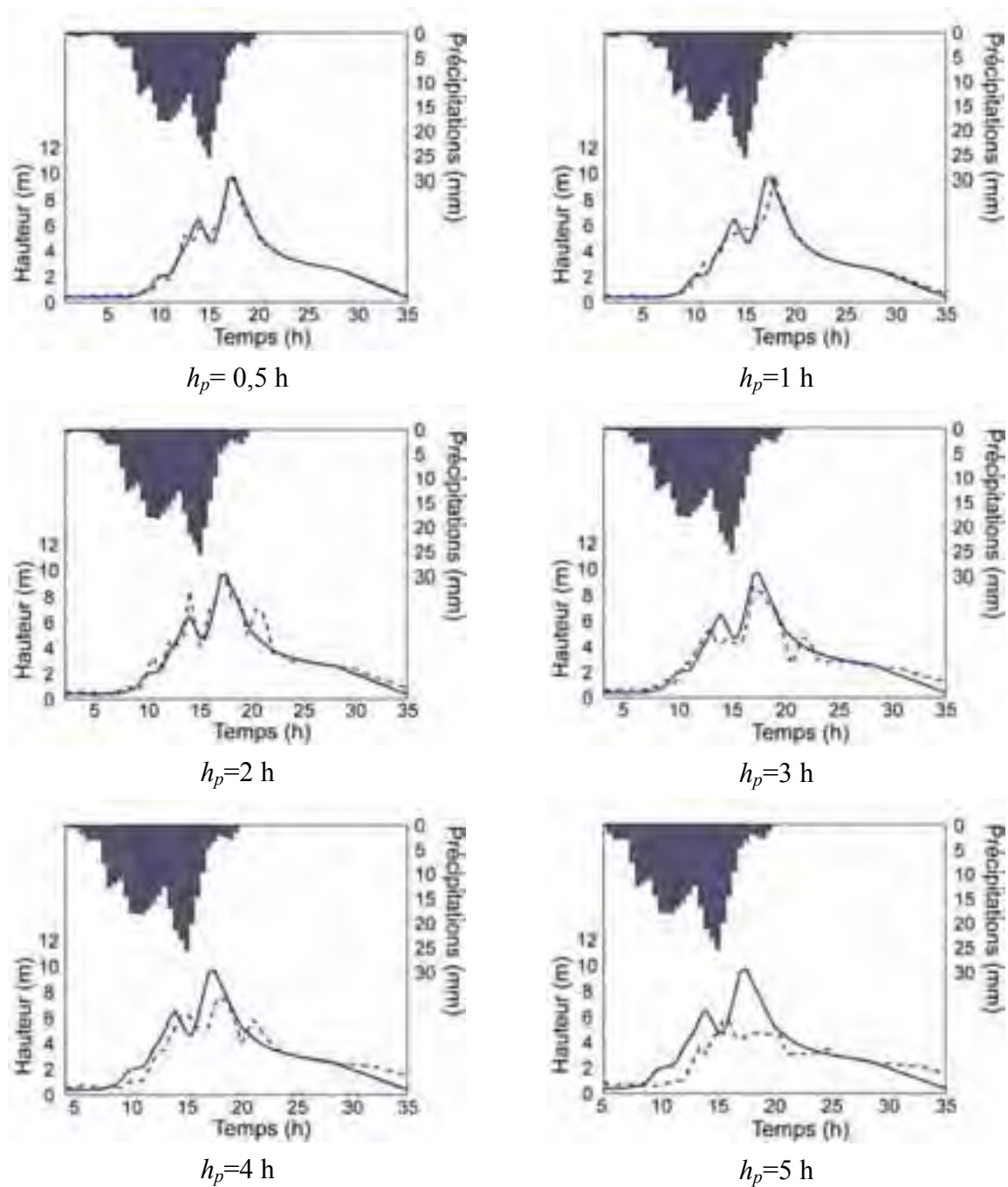


$h_p=5$  h

**Figure 23 : Meilleurs  $R^2$  calculés en fonction de la complexité**

L'événement de septembre 2002 est à la fois en arrêt et en test.





**Figure 24 : Limnigrammes de l'événement de 2002 utilisé en arrêt et en test**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	5	2	4	3	2	3	3
Critère de persistance $C_p$	0,67	0,68	0,83	0,86	0,81	0,66	0,75
$R^2$ (Critère de Nash)	0,99	0,95	0,93	0,91	0,82	0,54	0,86
Critère de hauteur $C_h$	0,94	0,87	0,90	0,82	0,76	0,60	0,82

**Tableau 7 : Critères de qualité de référence**

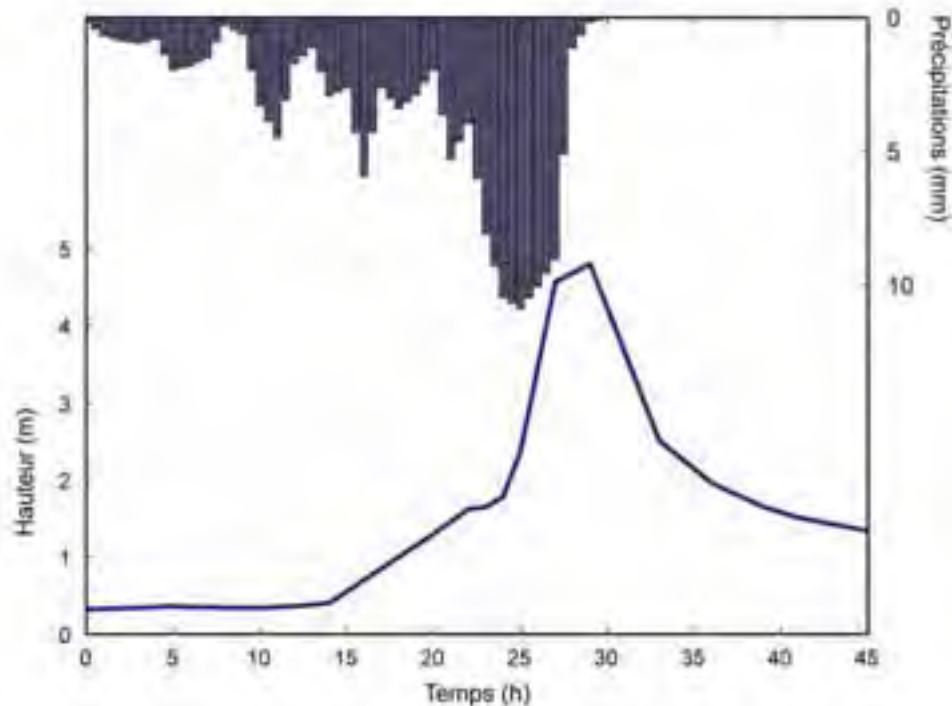
Ces limnigrammes et critères de qualité seront utilisés pour évaluer les différentes prévisions présentées dans la suite de ce chapitre.

#### L'ensemble d'arrêt est différent de l'événement de 2002

De la section précédente (III.5.1.c), nous pouvons conserver l'idée que l'ensemble d'arrêt « simple » conduit à un arrêt réellement précoce. Se pose alors la question de choisir pour l'ensemble d'arrêt un événement qui serait à la fois « simple », mais transcrirait le comportement de tous les autres.

Pour cela il a été postulé que cet événement pouvait être celui qui est le mieux prévu en test sur l'ensemble de la base. Pour cette application, il s'agit sans ambiguïté de l'événement 13 de septembre 2000, qui a donc été choisi comme ensemble d'arrêt.

Il est satisfaisant de noter que l'événement 13 est un événement simple, possédant un seul pic de crue culminant à 4,80 m et appartenant à la même catégorie que l'événement de 2002 (intense, mono pic, court) qui ne comporte que 2 représentants. Nous l'avons donc considéré comme un prototype de l'ensemble des événements disponibles. La Figure 25 représente le limnigramme de la crue de septembre 2000. L'arrêt précoce réalisé avec cet événement sera présenté au paragraphe III.5.1.f.



**Figure 25 : Précipitations et limnigramme de l'événement de septembre 2000.**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

#### *III.5.1.e Application de la validation croisée et de l'arrêt précoce avec un ensemble d'arrêt « simple »*

La procédure suivie est celle qui est décrite au paragraphe III.5.1.c lorsque les ensembles d'arrêt et de validation sont identiques. Néanmoins les fenêtres temporelles  $f_T$  et nombre de hauteurs mesurées appliquées ( $r$ ), ne sont pas réévaluées en raison de la complexité combinatoire que ceci implique ; ces grandeurs sont reportées pour mémoire dans le Tableau 8.

Notons que ni la crue majeure de septembre 2002, ni celle de septembre 2000, ne sont présentées aux différents réseaux lors de la phase de sélection du modèle.

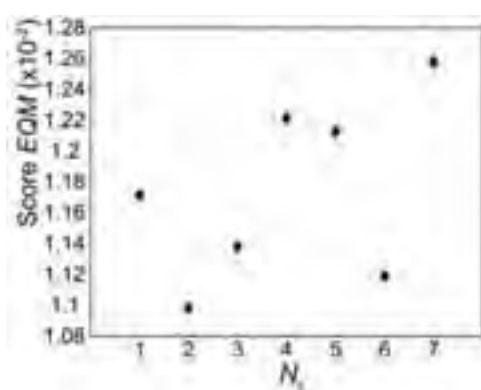
En comparant le Tableau 5 avec le Tableau 8, nous constatons que l'application de l'arrêt précoce entraîne une augmentation des écarts-types, quel que soit le critère de performance choisi : le modèle est encore plus sensible à l'initialisation des paramètres.

Par ailleurs, les scores optimaux du Tableau 8 sont la plupart du temps équivalents à ceux obtenus dans le Tableau 5.

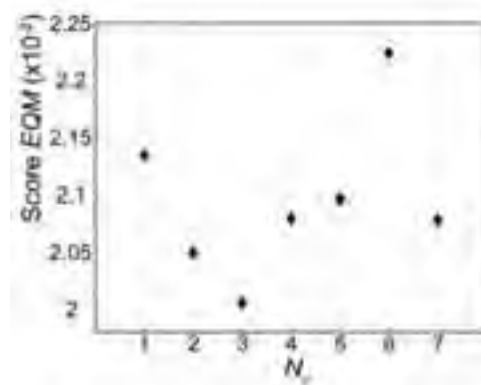
Il n'y a pas de nette amélioration apportée par l'application de l'arrêt précoce (réalisé ici à la 7<sup>ème</sup> itération : moyenne des itérations d'arrêt lors de la validation croisée).

Horizon de prévision $h_p$ (heures)		0,5	1	2	3	4	5
Fenêtre temporelle $f_T$ (heures)		2,5	3	3	2	0,5	0,5
Nombre de neurones cachés $N_c$		2	3	2	3	4	4
Retard $r$ sur $y^p$		2	2	2	2	2	2
Erreur Quadratique Moyenne EQM	Score minimum ( $\times 10^{-2}$ )	1,1	2,01	3,25	4,06	4,56	5,63
	Score moyen ( $\times 10^{-2}$ )	2,43	3,70	7,52	8,34	5,29	6,32
	ET (score) ( $\times 10^{-2}$ )	5,19	4,61	22,59	18,55	0,90	1,02
	Score médian ( $\times 10^{-2}$ )	1,45	2,64	3,64	4,68	5,10	6,13
Critère de hauteur $C_h$	Score maximum	0,95	0,92	0,88	0,85	0,81	0,76
	Score moyen	0,94	0,89	0,85	0,81	0,79	0,73
	ET (score) ( $\times 10^{-2}$ )	0,97	1,34	1,04	1,67	1,03	1,15
	Score médian	0,94	0,89	0,85	0,81	0,79	0,73

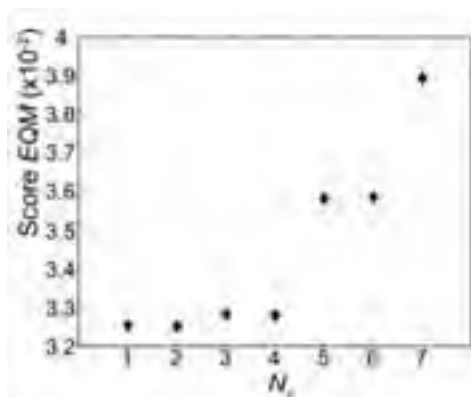
**Tableau 8 : Meilleurs modèles issus de la validation croisée avec arrêt précoce sur l'ensemble de validation**



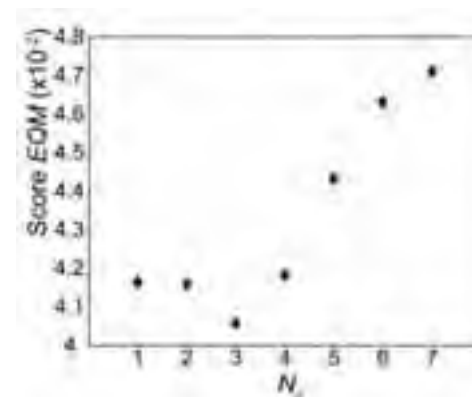
$h_p=0,5$  h



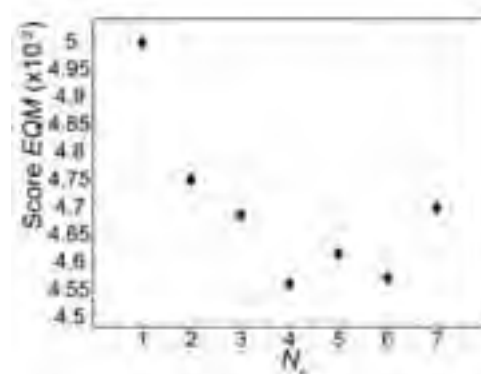
$h_p=1$  h



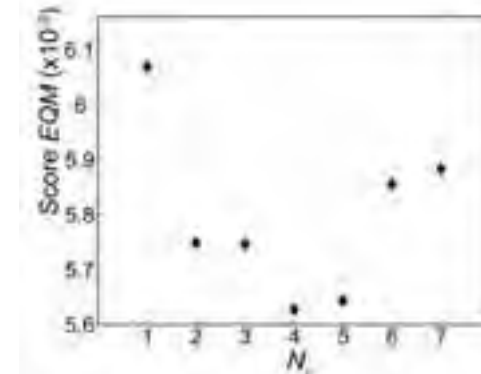
$h_p=2$  h



$h_p=3$  h



$h_p=4$  h



$h_p=5$  h

**Figure 26 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée avec arrêt précoce sur l'ensemble de validation, pour chaque horizon de prévision**

On peut relever, en résumé, que, quel que soit l'horizon de prévision, le nombre optimal de neurones de la couche cachée demeure faible (entre 2 et 4 neurones cachés). Les scores diminuent lorsque  $h_p$  augmente, ce qui traduit bien le fait que la prévision à horizon lointain est difficile à réaliser en l'absence de prévision de pluie.

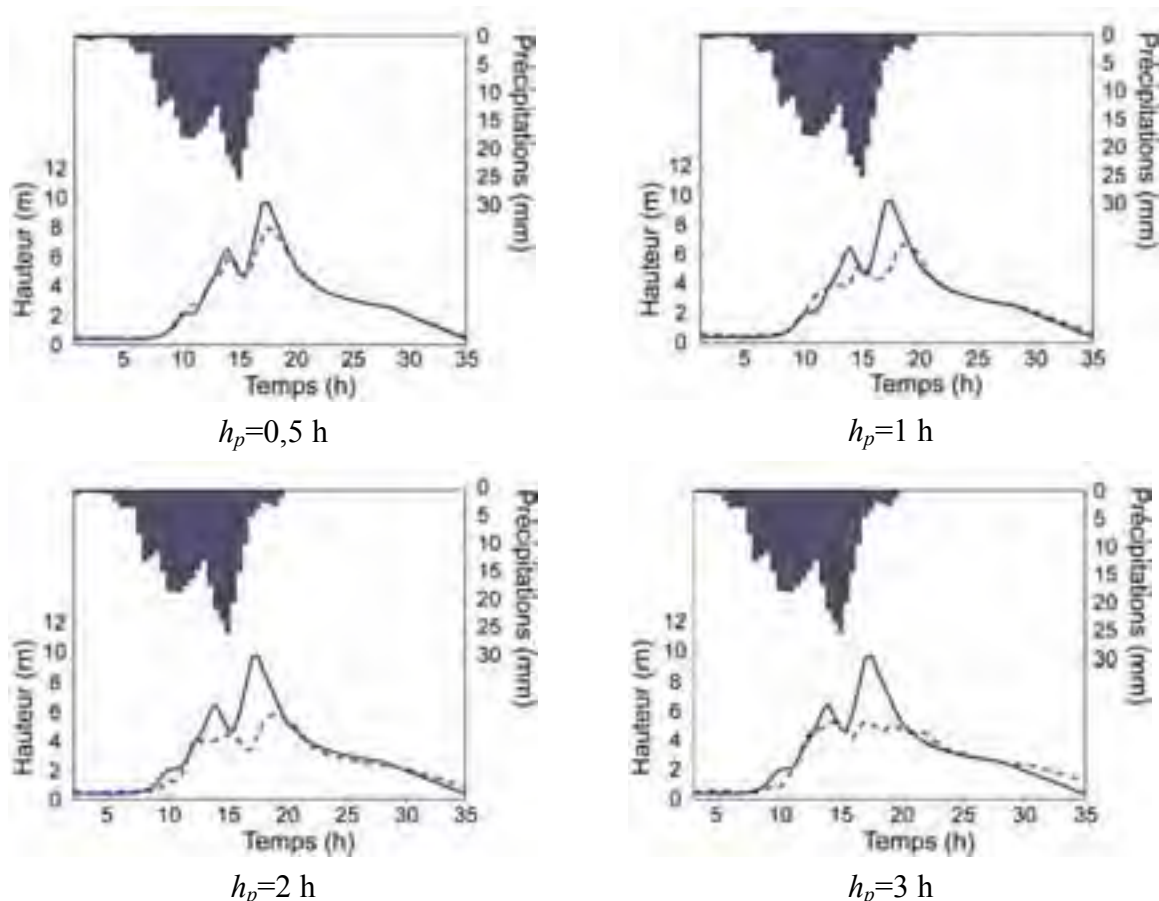
Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	3	2	3	4	4	3
Critère de persistance $C_p$	0,14	-0,30	0,26	0,63	0,74	0,54	0,34
$R^2$ (Critère de Nash)	0,96	0,79	0,68	0,76	0,76	0,39	0,72
Critère de hauteur $C_h$	0,86	0,67	0,59	0,72	0,68	0,54	0,68

**Tableau 9 : Critères de qualité des prévisions issues de la validation croisée avec arrêt précoce sur l'ensemble de validation, pour chaque horizon de prévision**

Avec l'application de l'arrêt précoce comme méthode de régularisation, nous remarquons une nette amélioration des trois critères de qualité. En effet, le critère de persistance moyen est passé de négatif (-0,28) à positif (0,34), le coefficient de détermination moyen est passé de 0,61 à 0,72, et le critère de hauteur moyen s'est aussi amélioré de 0,63 à 0,68.

Cependant l'analyse visuelle des limnigrammes est toujours décevante : hormis pour  $h_p=1/2$  heure, le pic de crue principal n'est pas, ou peu, reproduit.

Même si les indicateurs de qualité ont nettement progressé par l'introduction de l'arrêt précoce, les limnigrammes semblent indiquer, soit qu'une « saturation » par rapport à l'événement le plus intense de l'ensemble d'apprentissage a lieu, soit qu'il n'y a pas assez d'événements très intenses dans cet ensemble d'apprentissage. La validation croisée, qui par essence tient compte de tous les événements, converge vers une moyenne des comportements et ne parvient donc pas à sélectionner un modèle représentatif des comportement intenses.



**Figure 27 : Limnigrammes de test avec la validation croisée et l'arrêt précoce sur l'ensemble de validation.**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

### III.5.1.f Application de la validation croisée partielle

Partant de la dernière hypothèse émise au paragraphe précédent, nous proposons :

- de réaliser la sélection du modèle, non pas avec l'ensemble des événements, mais seulement à partir des événements intenses,
- d'utiliser l'événement 13 comme ensemble d'arrêt comme proposé au paragraphe III.5.1.d.

Ainsi, d'une part, le modèle est sélectionné pour les crues intenses, et, d'autre part, l'arrêt précoce est mis en œuvre avec un événement de type intense, qui représente le mieux possible l'ensemble de la base d'apprentissage.

Nous désignerons cette procédure sous le terme de *validation croisée partielle*.

Les événements les plus intenses utilisés pour calculer les scores, et sélectionner ainsi le modèle, sont déterminés par la typologie présentée au paragraphe II.5.3.c. Il s'agit des événements suivants :

- événement 03 (début octobre 1995),
- événement 04 (mi-octobre 1995),
- événement 109 (décembre 1997),
- événement 23 (octobre 2006).

La procédure fonctionne en deux étapes :

- sélection du modèle grâce à la validation croisée partielle ; pour chaque événement intense en validation, l'apprentissage est réalisé sur les 12 événements restants (les événements de 2002 et 2000 sont respectivement consacrés au test et à l'arrêt),
- une fois le modèle sélectionné pour chaque horizon de prévision, l'apprentissage est réalisé sur 13 événements (hormis ceux de septembre 2002 et septembre 2000).

La sélection du modèle conduit aux résultats présentés dans le Tableau 10.

En comparant les éléments du Tableau 10 avec ceux des Tableau 5 et Tableau 8 (validation croisée respectivement sans et avec arrêt précoce), il apparaît que, pour les faibles horizons de prévision, les écart-types diminuent, parfois notablement, en particulier pour les critères de persistance et de hauteur. Ceci traduit une meilleure convergence de la validation croisée partielle pour ces horizons. Pour les horizons supérieurs, les écarts-types sont assez peu affectés.

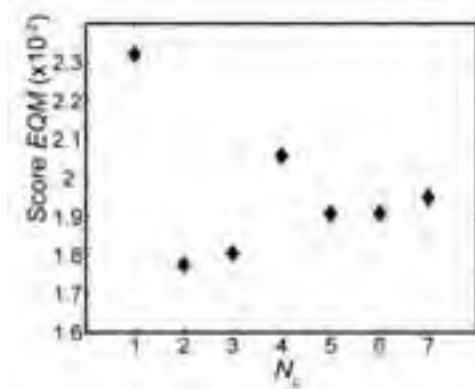
On peut noter également que la complexité évolue par rapport aux deux précédents tableaux, en particulier pour l'horizon de prévision de 2 h : le score calculé sur l'erreur quadratique (Figure 28) indique 7 neurones cachés, ce qui est surprenant car en général la complexité obtenue est assez faible. C'est pourquoi, pour ce dimensionnement particulier, nous avons préféré privilégier le score du critère de hauteur qui indique  $N_c=5$ .

Cependant, en général, les différents scores évoluent dans le même sens pour les trois critères et convergent vers les mêmes dimensionnements.

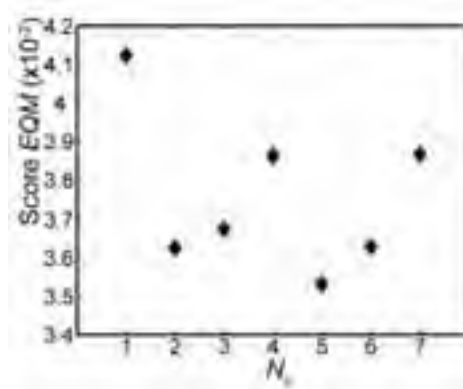


Horizon de prévision $h_p$ (heures)		0,5	1	2	3	4	5
Fenêtre temporelle $f_T$ (heures)		2,5	3	3	2	0,5	0,5
Nombre de neurones cachés $N_c$		2	2	5	3	3	3
Retard $r$ sur $y^p$		2	2	2	2	2	2
Initialisation		71	100	86	21	48	46
Erreur Quadratique Moyenne EQM	Score minimum ( $\times 10^{-2}$ )	1,78	3,63	5,17	5,50	6,94	9,27
	Score moyen ( $\times 10^{-2}$ )	2,81	4,84	7,05	7,53	8,85	10,99
	ET (score) ( $\times 10^{-2}$ )	0,65	0,69	0,87	0,81	1,04	1,02
	Score médian ( $\times 10^{-2}$ )	2,62	4,70	7,04	7,41	8,66	10,74
Critère de hauteur $C_h$	Score maximum	0,94	0,89	0,87	0,81	0,80	0,73
	Score moyen	0,91	0,85	0,79	0,76	0,75	0,66
	ET (score) ( $\times 10^{-2}$ )	1,09	1,70	2,27	2,05	2,15	2,66
	Score médian	0,91	0,85	0,78	0,76	0,75	0,66

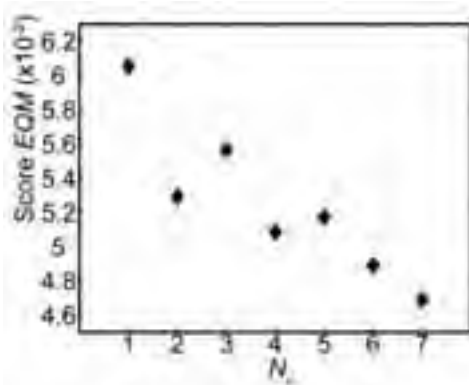
**Tableau 10 : Meilleurs modèles issus de la validation croisée partielle avec arrêt précoce sur l'événement 13**



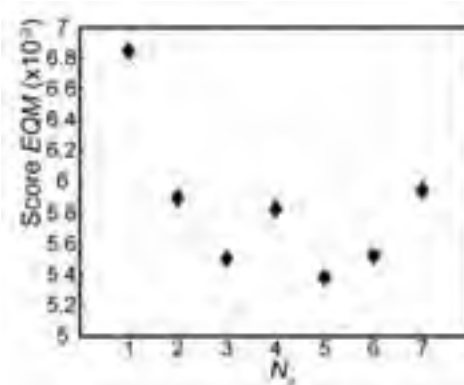
$h_p=0,5$  h



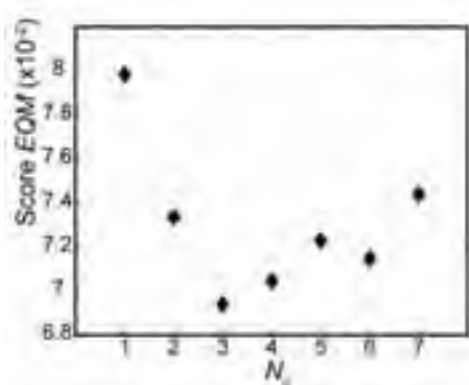
$h_p=1$  h



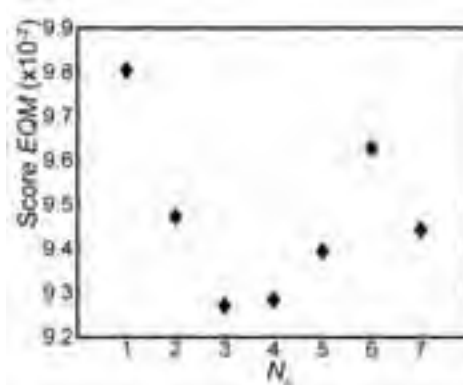
$h_p=2$  h



$h_p=3$  h



$h_p=4$  h



$h_p=5$  h

Figure 28 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec arrêt précoce sur l'événement 13, pour chaque horizon de prévision

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	2	5	3	3	3	3
Critère de persistance $C_p$	0,60	0,54	0,71	0,90	0,84	0,70	0,72
$R^2$ (Critère de Nash)	0,98	0,93	0,87	0,94	0,85	0,60	0,86
Critère de hauteur $C_h$	0,90	0,84	0,73	0,82	0,79	0,60	0,78

**Tableau 11 : Critères de qualité des prévisions issues de la validation croisée partielle avec arrêt précoce sur l'événement 13, pour chaque horizon de prévision**

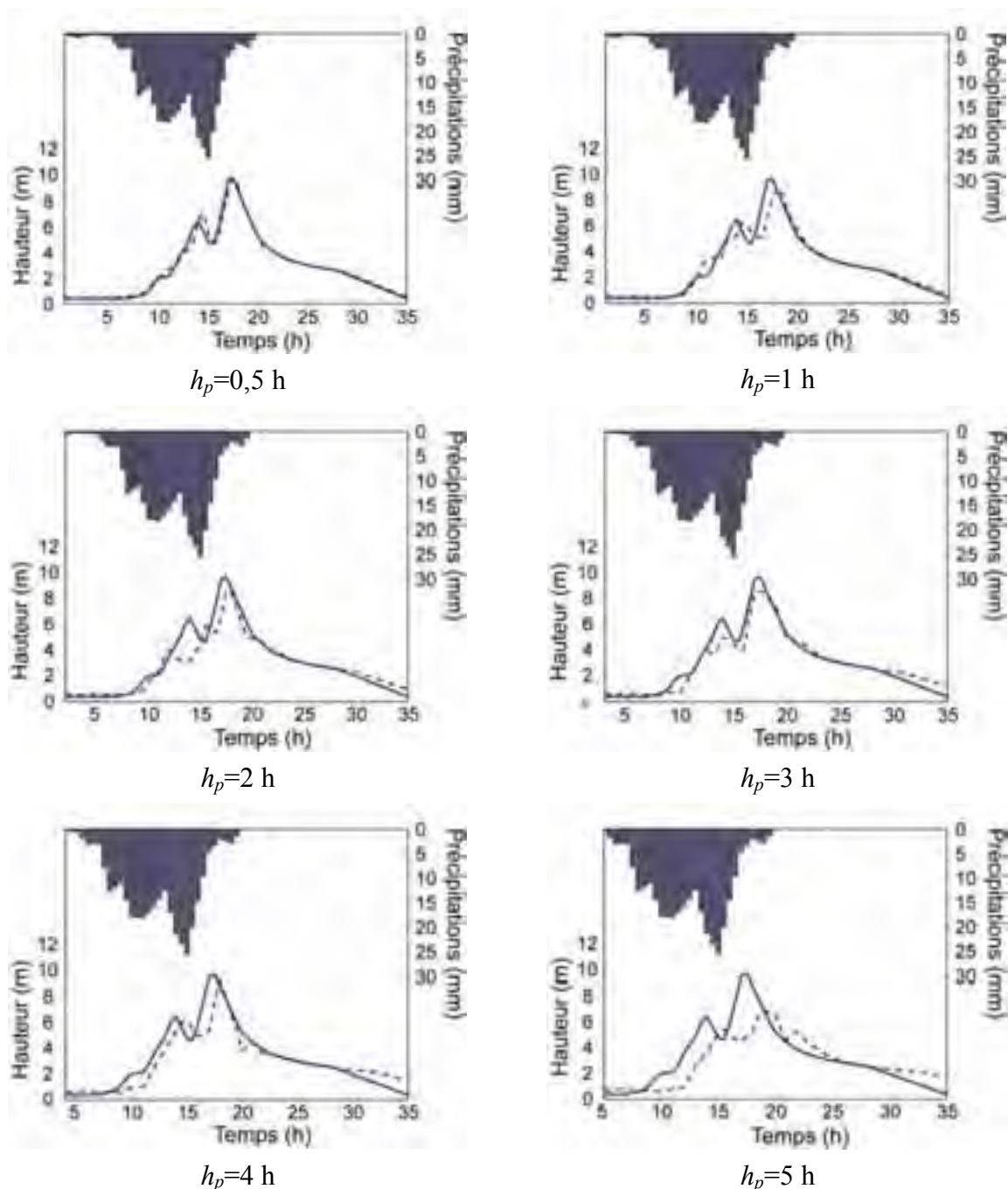
Quel que soit l'horizon de prévision  $h_p$ , le test sur l'événement de septembre 2002 donne lieu à des critères de persistance positifs.

Les coefficients de détermination  $R^2$  sont excellents pour les prévisions à court terme et vont en diminuant lorsque  $h_p$  augmente. Toutefois, nous devons mentionner la particularité de la prévision à 2 heures. Elle nécessite un nombre plus élevé de neurones cachés que pour les autres horizons de prévision, et si le score calculé avec l'erreur quadratique augmente lorsque l'horizon de prévision augmente (Tableau 10), en revanche, les critères de performances calculés sur l'ensemble de test (Tableau 11) ne suivent pas cette tendance pour l'horizon 2 h. Nous pensons que des erreurs de mesures particulières doivent affecter cet horizon pour quelques événements intenses dont celui de 2002. Ceci entraîne une augmentation de la complexité, soit 5 neurones cachés et une moins bonne prévision des pics de crue pour cet horizon que pour le suivant à 3 h (à  $h_p=2$  h,  $C_h = 0,73$  et  $R^2 = 0,87$ ).

En revanche, pour les prévisions à très court terme, soit à 30 mn ou à 1 h, les prévisions sont excellentes. Avec des critères de persistance nettement plus proches de 1, le prédicteur parvient enfin à faire bien mieux que la prévision naïve.

Si l'on souhaite comparer ces critères avec ceux obtenus par le test de référence (Tableau 7), on peut noter que les résultats obtenus sont tout à fait comparables pour les  $R^2$  et critère de hauteur, et qu'ils surpassent largement les critères de persistance précédemment obtenus : le prédicteur obtenu a donc d'excellentes propriétés d'anticipation.

L'analyse visuelle des limnigrammes de test sur l'événement de septembre 2002 est également excellente jusqu'à l'horizon de prévision 3 h, malgré le défaut signalé pour 2 h ; elle se dégrade un peu pour 4 h et est médiocre pour 5 h. On peut noter que, non seulement le pic de crue est bien estimé en hauteur, mais également qu'il est synchrone avec le pic mesuré pour les horizons de prévision jusqu'à 3 h. Pour 4 h, le retard du pic observé sur le pic observé n'est que de 1 h : il y a donc encore une anticipation de 3 h. À l'horizon de 5h, il n'y a pas à proprement parler de prévision de pic ; le limnigramme est cependant meilleur que celui de référence (Tableau 7).



**Figure 29 : Limnigrammes de test avec la validation croisée partielle et l'arrêt précoce sur l'événement 13**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Du point de vue hydrologique, ces résultats sont aussi très intéressants. Il faut en effet se rappeler que le temps de concentration du bassin versant est de quelques heures (2 h - 4 h) et que la prévision est réalisée en l'absence de prévision de pluie. En particulier pour l'événement très rapide de 2002, comme l'intensité la plus grande des pluies est tombée près d'Anduze et s'est donc trouvée en moins de 2 h dans la rivière, on demande au prédicteur de

prévoir la hauteur sans que la pluie (dont les prévisions ne sont pas disponibles) ne soit encore tombée. En particulier, la qualité de la prévision à l'horizon de 3 h est telle qu'elle laisse à penser que le prédicteur a pu extraire en interne une certaine « régularité » sur les précipitations et ainsi l'utiliser pour estimer la hauteur d'eau adéquate.

Enfin, le fait que ces résultats soient du même ordre de grandeur ou meilleurs que ceux obtenus comme « référence » est réellement satisfaisant : la validation croisée partielle réussit vraisemblablement à tirer le meilleur parti de l'architecture et des variables utilisées.

### III.5.2 Régularisation par modération de poids

#### III.5.2.a Présentation

Après avoir appliqué avec succès l'arrêt précoce, méthode active de régularisation, il paraît intéressant d'étudier la régularisation par modération de poids. Cette dernière a été présentée au chapitre I ; rappelons qu'elle contraint les paramètres du réseau à ne pas prendre de valeurs excessivement grandes.

Ainsi dans le cadre de cette étude, la nouvelle fonction de coût  $J'$  a l'expression suivante :

$$J' = J + \frac{1-\gamma}{2} \|\theta\|^2$$

où :

- $J$  est l'erreur quadratique moyenne habituellement utilisée,
- $\theta$  est le vecteur des paramètres du réseau,
- $\gamma$  est l'hyperparamètre qui détermine l'importance relative des deux termes dans la fonction de coût  $J'$ .

#### III.5.2.b Sélection du modèle par validation croisée partielle

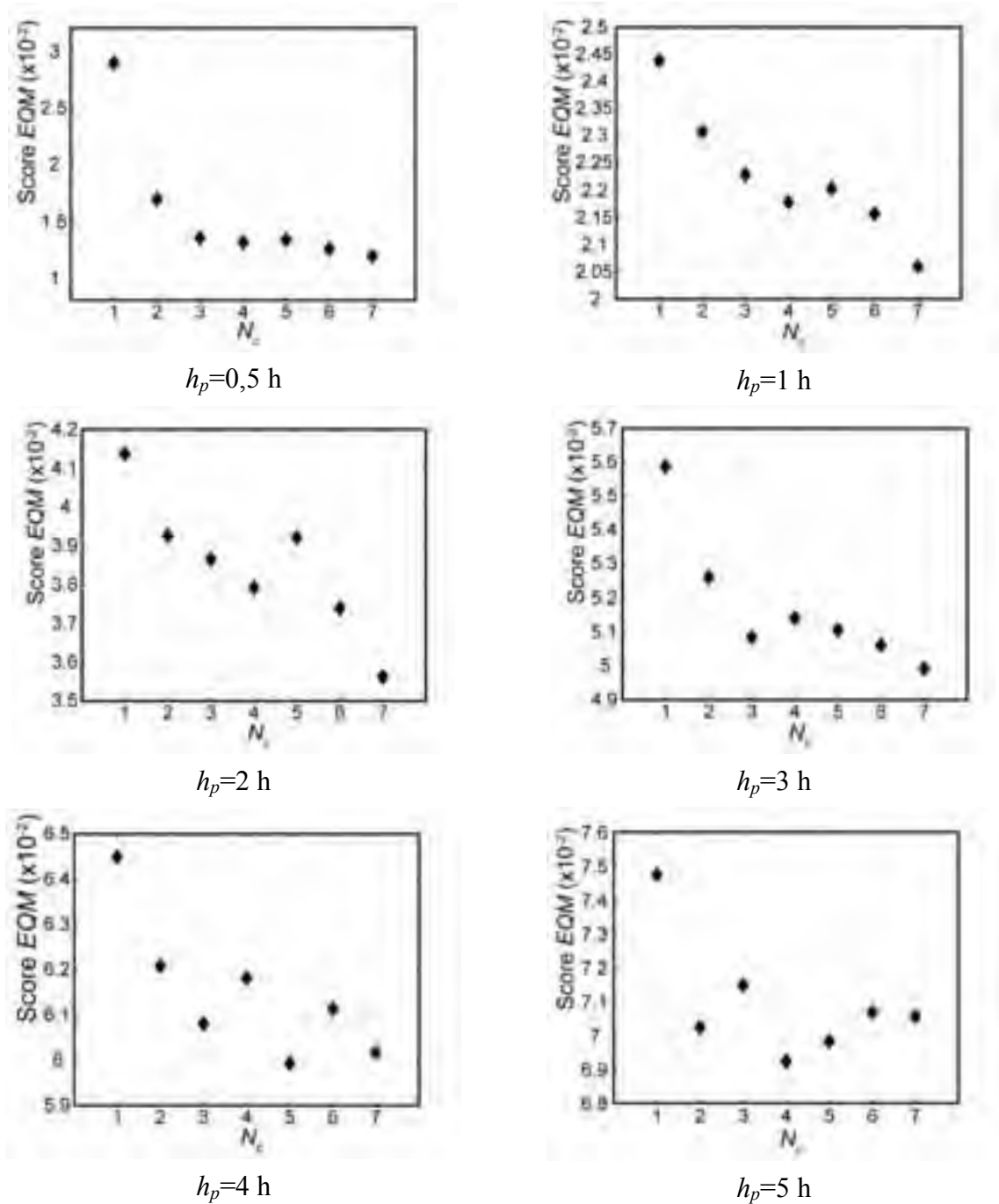
Nous sommes amenés à recalculer la complexité du réseau en appliquant la méthode de régularisation par modération de poids. Tout comme lors de la sélection du modèle, l'hyperparamètre  $\gamma$  est sélectionné par validation croisée partielle, pour chaque horizon de prévision  $h_p$ , pour des valeurs de  $\gamma$  allant de 0,5 à 0,95 avec un incrément de 0,05.

Toujours pour éviter une explosion combinatoire, la sélection du modèle a été réalisée en considérant les valeurs de la fenêtre temporelle et du nombre d'entrée des hauteurs d'eau passées à l'identique par rapport à la première sélection.

L'analyse de ce tableau fait apparaître que l'application de la modération de poids entraîne une augmentation du nombre de neurones cachés pour tous les horizons de prévision. On note également que les scores d'erreur quadratique moyenne minimums sont légèrement plus faibles que ceux issus de l'application de la régularisation par arrêt précoce. De plus, la régularisation par modération des poids donne lieu à des écarts-types plus faibles. Les solutions trouvées par cette méthode de régularisation sont donc moins dépendantes de l'initialisation des paramètres que celles issues de l'application de l'arrêt précoce.

Horizon de prévision $h_p$		0,5	1	2	3	4	5
$\gamma$		0,60	0,90	0,70	0,55	0,75	0,75
Nombre de neurones cachés $N_c$		7	7	7	7	5	4
Erreur Quadratique Moyenne EQM	Score minimum ( $\times 10^{-2}$ )	1,19	2,06	3,56	4,99	5,99	6,92
	Score moyen ( $\times 10^{-2}$ )	1,50	2,57	4,09	5,39	6,58	7,60
	ET (score) ( $\times 10^{-2}$ )	0,19	0,19	0,14	0,15	0,27	0,41
	Score médian ( $\times 10^{-2}$ )	1,45	2,54	4,08	5,39	6,53	7,51
Critère de hauteur $C_h$	Score maximum	0,95	0,92	0,86	0,81	0,80	0,72
	Score moyen	0,93	0,88	0,83	0,79	0,77	0,69
	ET (score) ( $\times 10^{-2}$ )	1,05	1,45	0,84	0,89	2,14	2,29
	Score médian	0,93	0,89	0,83	0,80	0,77	0,69

**Tableau 12 : Meilleurs modèles issus de la validation croisée partielle avec la modération des poids**



**Figure 30 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec modération des poids**

L'examen des évolutions des scores, représentés pour l'erreur quadratique moyenne sur la Figure 30, révèle des évolutions sensiblement différentes. Si l'arrêt précoce permettait d'identifier une complexité assez faible avec un minimum bien marqué, en revanche l'application de la modération des poids conduit à des scores qui décroissent puis se stabilisent quand la complexité augmente. Sachant que, à partir d'un certain seuil, plus la complexité augmente et plus la généralisation est difficile, il ne nous a pas paru raisonnable d'augmenter le nombre de neurones cachés au-delà de 7 ; c'est pourquoi cette valeur est celle qui est souvent choisie.

Par ailleurs, au regard du Tableau 13, en moyenne, les trois critères de qualité sont moins bons que ceux issus de l'arrêt précoce.

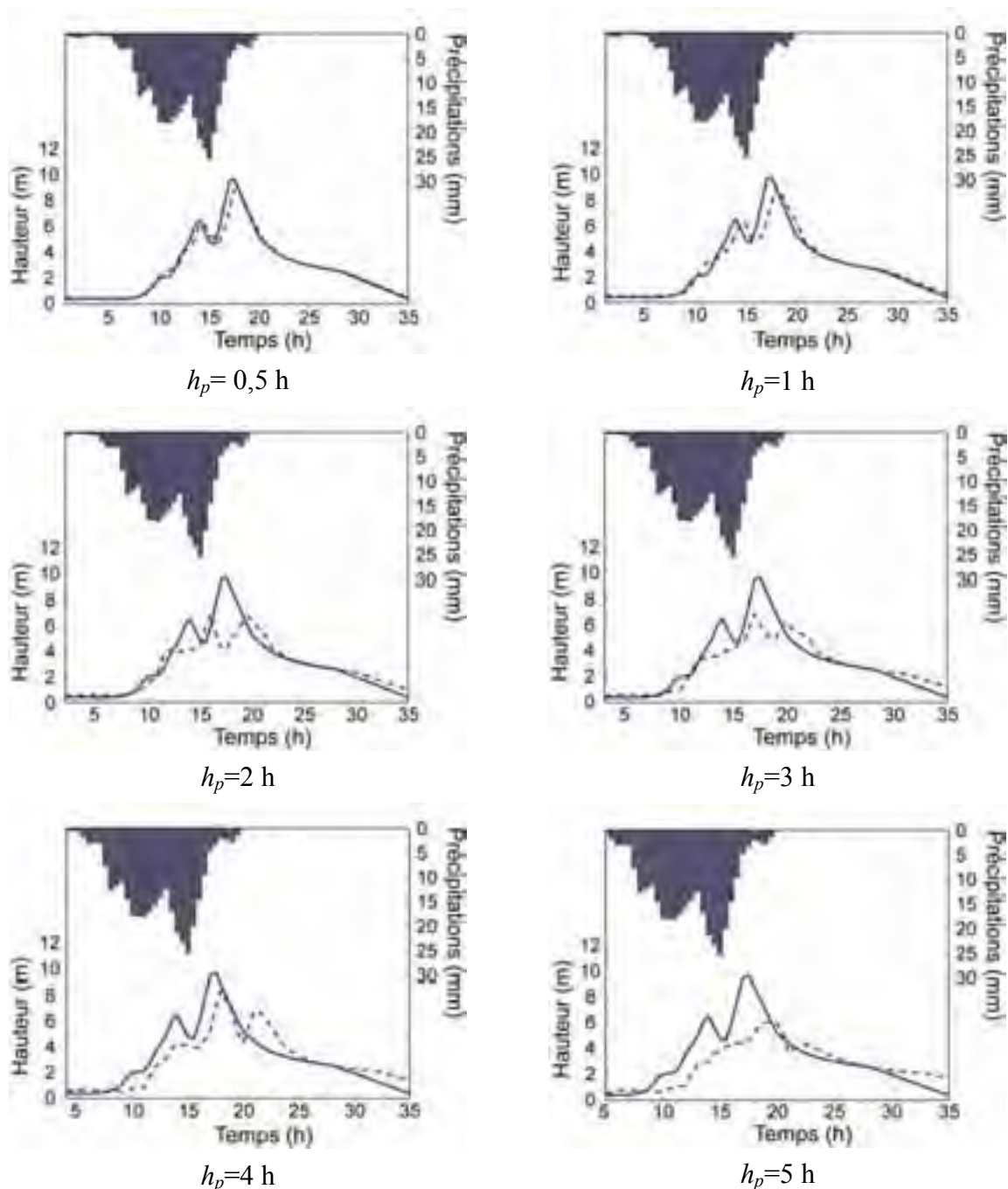
On peut également noter que le critère de persistance croît lorsque l'horizon de prévision augmente ; ceci ne signifie pas que la prévision s'améliore (dans le même temps le  $R^2$  diminue), mais simplement que pour les grands horizons, la prévision naïve n'a plus grand sens : il est donc relativement facile de réaliser une meilleure estimation, même si cette dernière est médiocre. Le critère de persistance a donc surtout un sens pour les horizons faibles ou moyens ; néanmoins, même en ne considérant que ces derniers, les valeurs du critère de persistance sont positives avec la modération de poids mais plus faibles que celles obtenues avec l'arrêt précoce.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	7	7	7	7	5	4	6
Critère de persistance $C_p$	0,23	0,47	0,34	0,68	0,71	0,68	0,52
$R^2$ (Critère de Nash)	0,97	0,92	0,72	0,79	0,73	0,58	0,79
Critère de hauteur $C_h$	0,86	0,81	0,65	0,69	0,69	0,55	0,71

**Tableau 13 : Critères de qualité des prévisions issues de la validation croisée partielle avec modération des poids**

L'analyse visuelle des limnigrammes de test sur l'événement de 2002 présentés sur la Figure 31, laisse apparaître que les limnigrammes ne sont pas à la hauteur des valeurs données par le critère de persistance. En effet si les deux limnigrammes obtenus pour les horizons 1 h et 5 h sont relativement semblables à ceux obtenus précédemment ; en revanche, pour les autres horizons, le pic de crue est bien mieux estimé, et de manière plus synchrone, par l'arrêt précoce.





**Figure 31 : Limnigrammes de test obtenus avec la validation croisée partielle et la régularisation par modération des poids**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Ainsi, la régularisation par modération des poids ne permet pas d'améliorer le critère de persistance sur l'ensemble de l'événement, par rapport à ce qui a été obtenu avec l'arrêt précoce. Elle ne permet donc pas de prévoir correctement les pics de crue pour le problème que nous avons étudié.

### III.5.3 Régularisation par cumul de l'arrêt précoce et de la modération de poids

#### III.5.3.a Sélection du modèle par validation croisée partielle

Afin d'évaluer si les avantages des deux méthodes de régularisation utilisées (modération des poids et arrêt précoce) peuvent se cumuler, nous proposons d'appliquer ici les deux méthodes conjointement. Le modèle est toujours sélectionné avec la validation croisée partielle qui permet de sélectionner les comportements intenses.

L'application simultanée de l'arrêt précoce et de la modération de poids se fait en arrêtant la phase d'apprentissage au minimum de la fonction de coût  $J'$  calculée sur l'ensemble d'arrêt. Cette procédure sera appelée *double régularisation*.

Les ensembles d'apprentissage, d'arrêt, de validation et de test sont choisis comme précédemment.

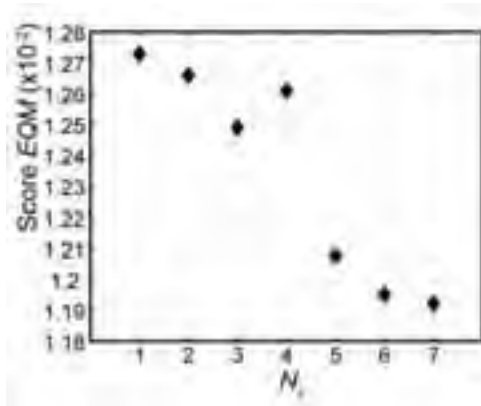
L'hyperparamètre  $\gamma$  varie également entre 0,5 et 0,95 avec un incrément de 0,05. Il est intéressant de noter sur le tableau récapitulatif, Tableau 14, que pour les faibles horizons de prévision, la fonction de coût privilégiée par  $\gamma$  est l'erreur quadratique moyenne, tandis que pour les horizons de 4 h et 5 h, les deux termes de la fonction de coût interviennent à hauteur équivalente.

Sur le Tableau 14, on peut noter que les modèles donnent des résultats de qualité assez semblable à ceux présentés dans le Tableau 12 (par la seule modération des poids) hormis une diminution de la complexité pour les horizons de 4 h et 5 h. La double régularisation entraîne également une légère augmentation des écarts-types par rapport à la modération des poids seule. C'est donc l'arrêt précoce qui est capable de conduire les paramètres du réseau à des solutions très diverses.

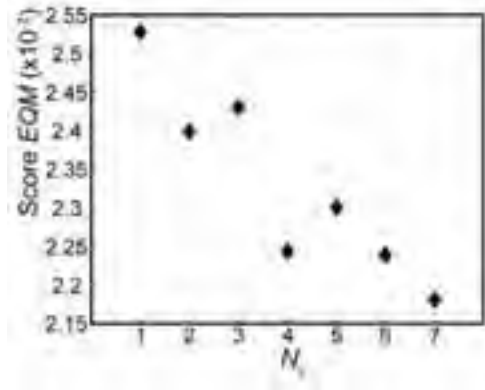
Du point de vue des évolutions des scores, on peut noter sur la Figure 32 qu'ils décroissent globalement lorsque la complexité augmente. Dans certains cas, comme pour l'horizon de 3 h, un net minimum apparaît. L'arrêt précoce tend donc à diminuer la complexité du réseau.

Horizon de prévision $h_p$		0,5	1	2	3	4	5
$\gamma$		0,85	0,85	0,90	0,85	0,50	0,55
Nombre de neurones cachés $N_c$		7	7	7	5	6	6
Erreur Quadratique Moyenne EQM	Score minimum ( $\times 10^{-2}$ )	1,19	2,18	3,56	4,92	6,15	7,06
	Score moyen ( $\times 10^{-2}$ )	1,48	2,89	4,18	5,58	8,89	10,53
	ET (score) ( $\times 10^{-2}$ )	0,33	1,24	0,29	0,76	2,75	3,23
	Score médian ( $\times 10^{-2}$ )	1,40	2,58	4,15	5,40	7,68	9,31
Critère de hauteur $C_h$	Score maximum	0,95	0,91	0,86	0,81	0,80	0,73
	Score moyen	0,93	0,88	0,82	0,78	0,68	0,60
	ET (score) ( $\times 10^{-2}$ )	1,14	2,26	1,39	2,04	8,70	8,18
	Score médian	0,93	0,89	0,83	0,78	0,70	0,61

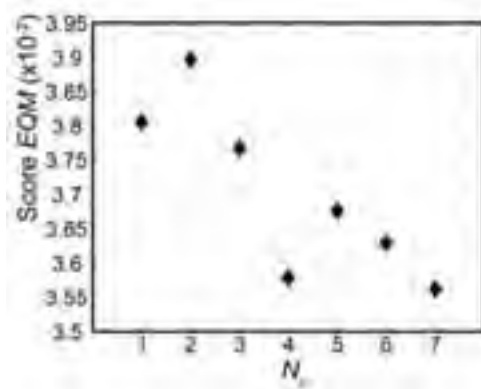
**Tableau 14 : Meilleurs modèles issus de la validation croisée partielle avec la double régularisation**



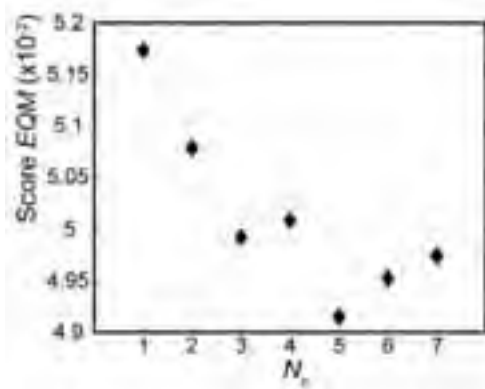
$h_p=0.5$  h



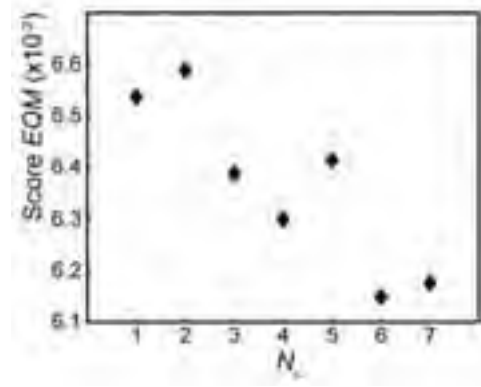
$h_p=1$  h



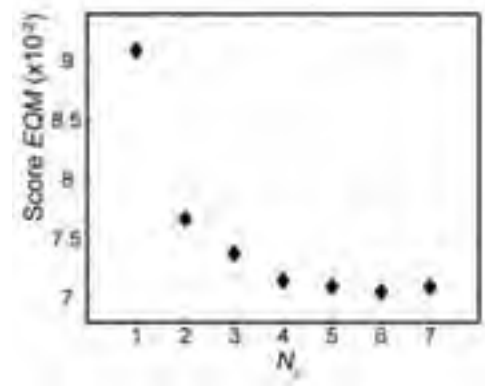
$h_p=2$  h



$h_p=3$  h



$h_p=4$  h



$h_p=5$  h

Figure 32 : Evolution des scores calculés avec l'erreur quadratique moyenne, issus de la validation croisée partielle avec double régularisation.

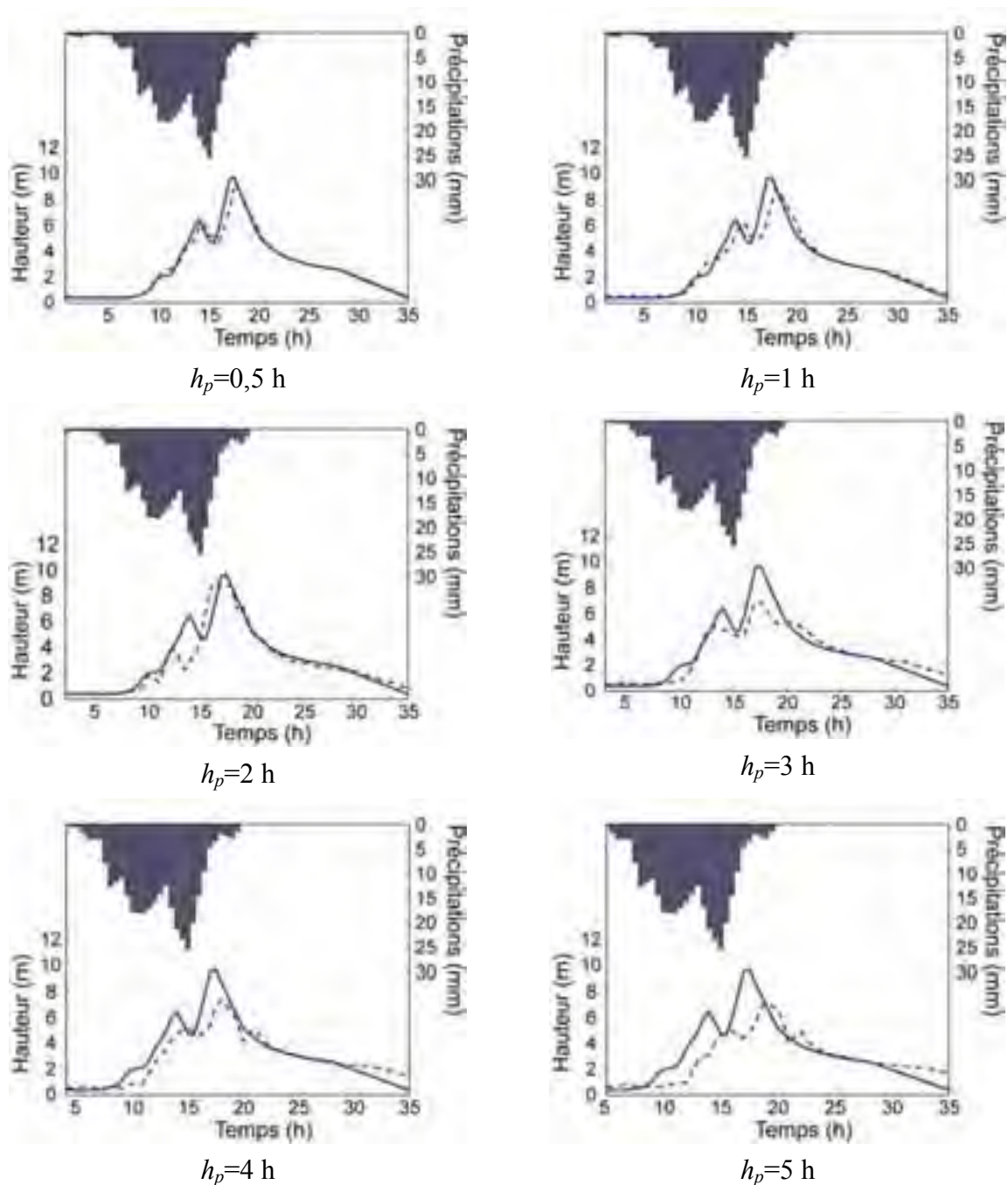
Au niveau des critères de qualité reportés dans le Tableau 15, tous sont améliorés avec la double régularisation par rapport à l'application de la modulation de poids seule ; néanmoins, ils restent moins satisfaisants qu'avec arrêt précoce seul. De même, l'analyse comparée des

Tableau 13, Tableau 11 et Tableau 15 permet de supposer que nous sommes en position intermédiaire entre les deux méthodes : il ne semble pas que l'on ait pu tirer profit simultanément des qualités des deux méthodes mises en œuvre individuellement.

Enfin l'analyse visuelle des limnigrammes de test sur l'événement de 2002 (Figure 33) confirme les hypothèses précédentes : les limnigrammes sont un peu meilleurs que ceux de la régularisation par modération des poids et moins bons que ceux obtenus par arrêt précoce. On doit cependant noter une exception pour l'horizon 2 h qui présente un limnigramme en avance sur la crue ; même si les critères qui en résultent sont médiocres, l'intérêt pour le prévisionniste est évident.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	7	7	7	5	6	6	6
Critères de Persistance $C_p$	0,23	0,47	0,68	0,79	0,81	0,70	0,61
$R^2$ (Critère de Nash)	0,97	0,92	0,86	0,86	0,82	0,61	0,84
Critère de hauteur $C_h$	0,86	0,81	0,71	0,78	0,73	0,59	0,75

**Tableau 15 : Critères de qualité des prévisions issues de la validation croisée avec double régularisation, pour chaque horizon de prévision.**



**Figure 33 : Limnigrammes de test avec la validation croisée partielle et la double régularisation.**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

### III.5.4 Synthèse de l'application des méthodes de régularisation

#### III.5.4.a Résultats de prévision

La meilleure prévision à laquelle nous avons abouti est celle issue de l'application de la validation croisée partielle avec un arrêt précoce sur la crue de septembre 2000.

C'est la méthode de prévision la plus parcimonieuse, qui ne nécessite en moyenne que 3 neurones cachés, et qui fournit les meilleures prévisions sur la crue de septembre 2002. Le critère de persistance moyen est 0,72 : la prévision est donc largement meilleure que la prévision naïve, mais moins bonne que la prévision de référence ( $C_p = 0,75$ ). Le coefficient de détermination  $R^2$  moyen vaut 0,86 comme pour la prévision de référence, ce qui est excellent ; enfin, le critère de hauteur moyen sur les différents horizons de prévision est égal à 0,78 contre 0,82 pour les prévisions de référence. À notre connaissance, aucun autre modèle de prévision n'a pu obtenir de tels résultats, en l'absence de prévision de pluie.

Par ailleurs, il convient de revenir sur un choix qui a été réalisé lors de la mise en œuvre de la validation croisée partielle : le modèle est sélectionné par rapport aux événements les plus intenses. Le risque est que les événements peu intenses, en l'absence de prévision de pluie soient considérés à tort comme des événements intenses et donnent lieu à des prévisions de pics de crue surestimées, voire complètement erronées.

Il est donc nécessaire de valider la méthode en effectuant un test sur d'autres événements que celui de 2002, moyennement intenses ou peu intenses. Ceci a été réalisé sur tous les événements de la base (Annexe G), mais n'est présenté ici que pour les 3 événements ci-après :

- événement 4 d'octobre 1995, intense, long avec un seul pic,
- événement 107 de novembre 1997, peu intense, long, avec 3 pics,
- événement 23 d'octobre 2006, intense, long avec 1 seul pic.

La procédure suivante a été suivie : avec le modèle sélectionné, pour chaque événement évalué, un nouvel apprentissage est réalisé pour chacun des horizons de prévision, avec tous les événements de la base, sauf l'événement 13 utilisé en ensemble d'arrêt, et l'événement sur lequel a lieu le test.

Le modèle utilisé, pour chaque horizon de prévision, est le modèle qui a été sélectionné lors de la validation croisée partielle exposée à la section III.5.1.f. En effet, du fait de la complexité combinatoire que la sélection complète du modèle engendre, les dimensionnements des modèles n'ont pas été refaits. Il faut préciser que si cette procédure est rigoureusement indépendante de l'ensemble de test pour l'événement 107 ainsi que pour tous les événements non intenses, elle ne l'est pas totalement pour les autres événements, qui appartiennent à la catégorie des événements intenses et ont donc servi à réaliser la sélection du modèle utilisé. Nous considérons cependant que, en pratique, cette réserve peut être écartée, car les scores de validation des événements lors de la validation croisée partielle sont en général tous minimums pour la même initialisation : un seul événement ne joue donc pas un grand rôle dans la décision. Cette dernière est donc la même, que l'événement de test participe ou pas à la sélection du modèle.

Les limnigrammes et les critères de qualité sont présentés ci-après et, avant d'entrer dans le détail, il apparaît que les résultats sont également assez satisfaisants.

### **Événement 4 d'octobre 1995**

Les limnigrammes de test réalisés sur l'événement 4 sont présentés sur la Figure 34 ; on peut remarquer que, si les résultats sont corrects jusqu'à l'horizon 2 h, ils se détériorent ensuite.

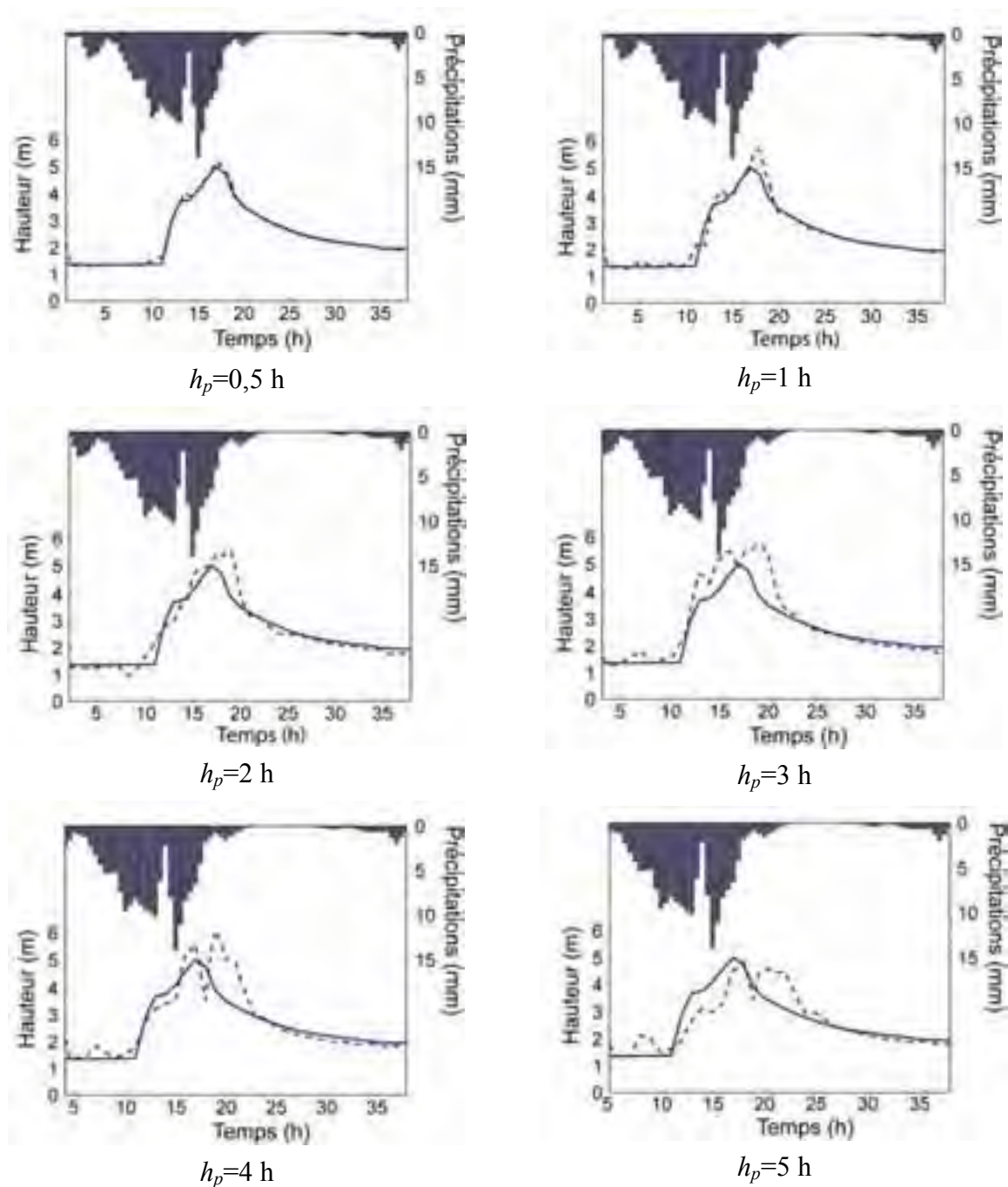
Principalement le réseau réalise une surestimation du pic dès l'horizon 1 h. Ce résultat est surprenant car le défaut de tous les prédicteurs est généralement de sous-estimer le pic de crue. Nous sommes donc en présence d'un comportement atypique que l'on pourrait expliquer en revenant sur le mode d'acquisition des signaux. Il faut se rappeler que les événements antérieurs à l'année 2002 sont échantillonnés toutes les heures, et que nous avons montré que cette période est trop longue. Il est donc vraisemblable que cet événement n'ait pas un maximum de crue à 864 m<sup>3</sup>/s, mais un pic bien plus élevé <sup>15</sup>! C'est bien ce que prévoit le prédicteur, même avec l'information de hauteur peut-être sous-estimée en entrée. Avec l'acquisition de nouveaux événements, il faudra peu à peu éliminer ces événements intenses anciens, mal échantillonnés, de la base de données.

L'analyse des critères de qualité met en évidence que les trois critères sont bons. La prévision de cet événement est donc largement meilleure que la prévision naïve, en dépit de la surestimation du pic.

---

<sup>15</sup> En fait, nous disposons d'un autre enregistrement de la crue de l'événement 4 à la période de 12 mn, qui donne le maximum du pic de crue à 1411 m<sup>3</sup>/s au lieu de 864 m<sup>3</sup>/s sur la base de données BVNE. Cet écart peut être expliqué aussi bien par le choix d'une courbe de tarage différente que par le ré-échantillonnage à 1 h des événements anciens ; malheureusement, l'origine du fichier échantillonné à 12 mn et des transformations subies ont été perdues. Il reste cependant que l'on peut soupçonner que l'amplitude du pic ait pu être bien plus élevée que celle conservée dans la base de données BVNE.





**Figure 34 : Limnigrammes de test sur l'événement 4 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

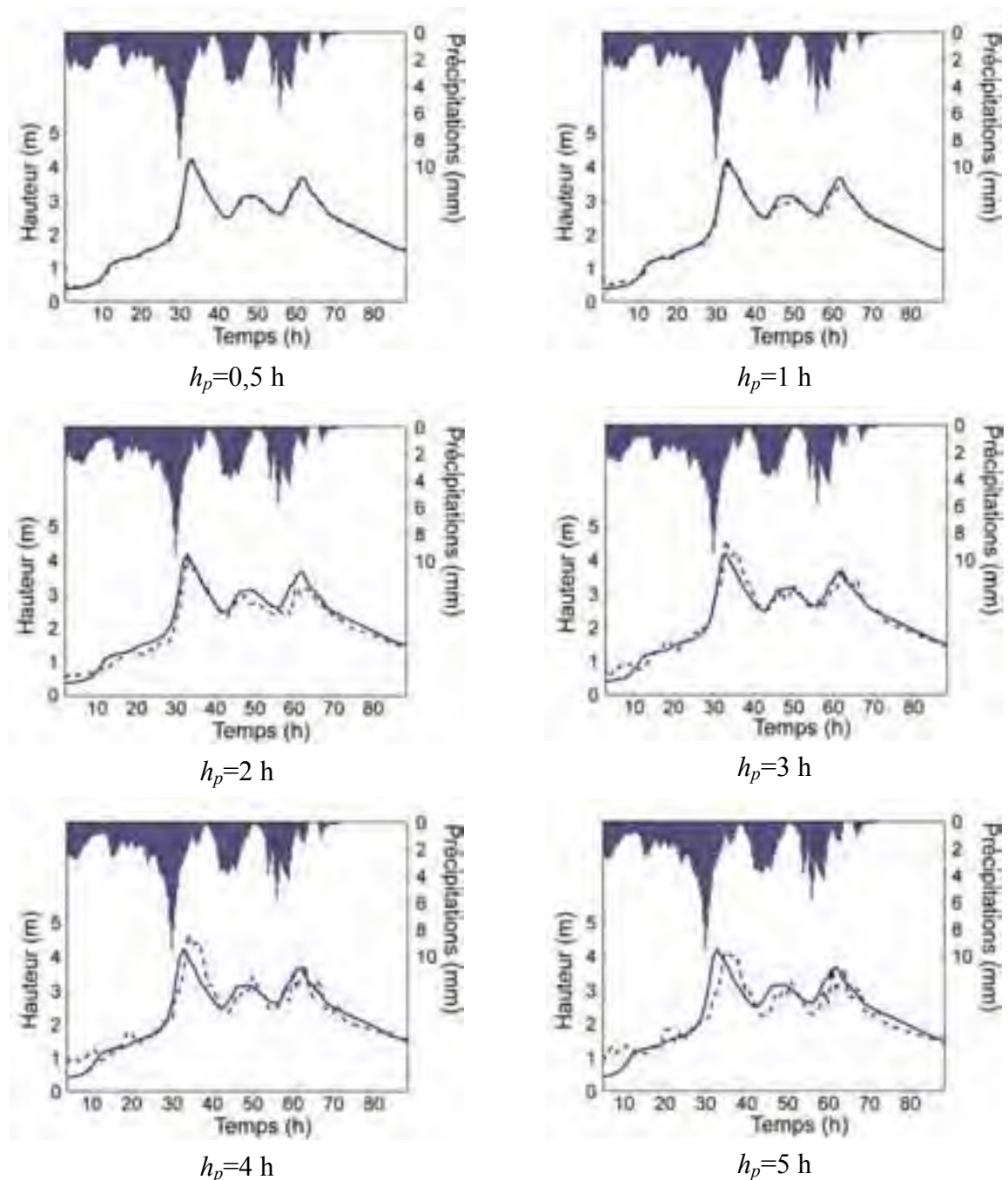
Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	2	5	3	3	3	2
Critère de persistance $C_p$	0,48	0,44	0,60	0,51	0,66	0,78	0,58
$R^2$ (Critère de Nash)	0,99	0,94	0,87	0,70	0,68	0,69	0,81
Critère de hauteur $C_h$	0,98	0,96	0,94	0,88	0,92	0,83	0,92

**Tableau 16 : Critères de qualité des prévisions sur l'événement 04**

### **Événement 107 de novembre 1997.**

Les prévisions effectuées sur l'événement 107 sont présentées car il correspond à la catégorie des événements complexes qui comprennent plusieurs passages orageux et pour lesquels la prise en considération de l'humidité des sols est critique. Il est donc intéressant de l'évaluer en test et de vérifier que le prédicteur ne le surestime pas. La difficulté liée à l'échantillonnage soulevée précédemment n'est pas critique ici, car cet événement est assez lent : l'échantillonnage à 1 h ne conduit pas à une modification des pics de crue.

L'analyse des critères de qualité (Tableau 17) met en évidence le même type de tendances que pour l'événement 4 : les critères de qualité sont excellents. Le critère de hauteur moyen avoisine l'unité, ce qui est excellent si l'on n'observe pas de pics prévus en dehors des pics principaux.



**Figure 35 : Limnigrammes de test sur l'événement 4 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

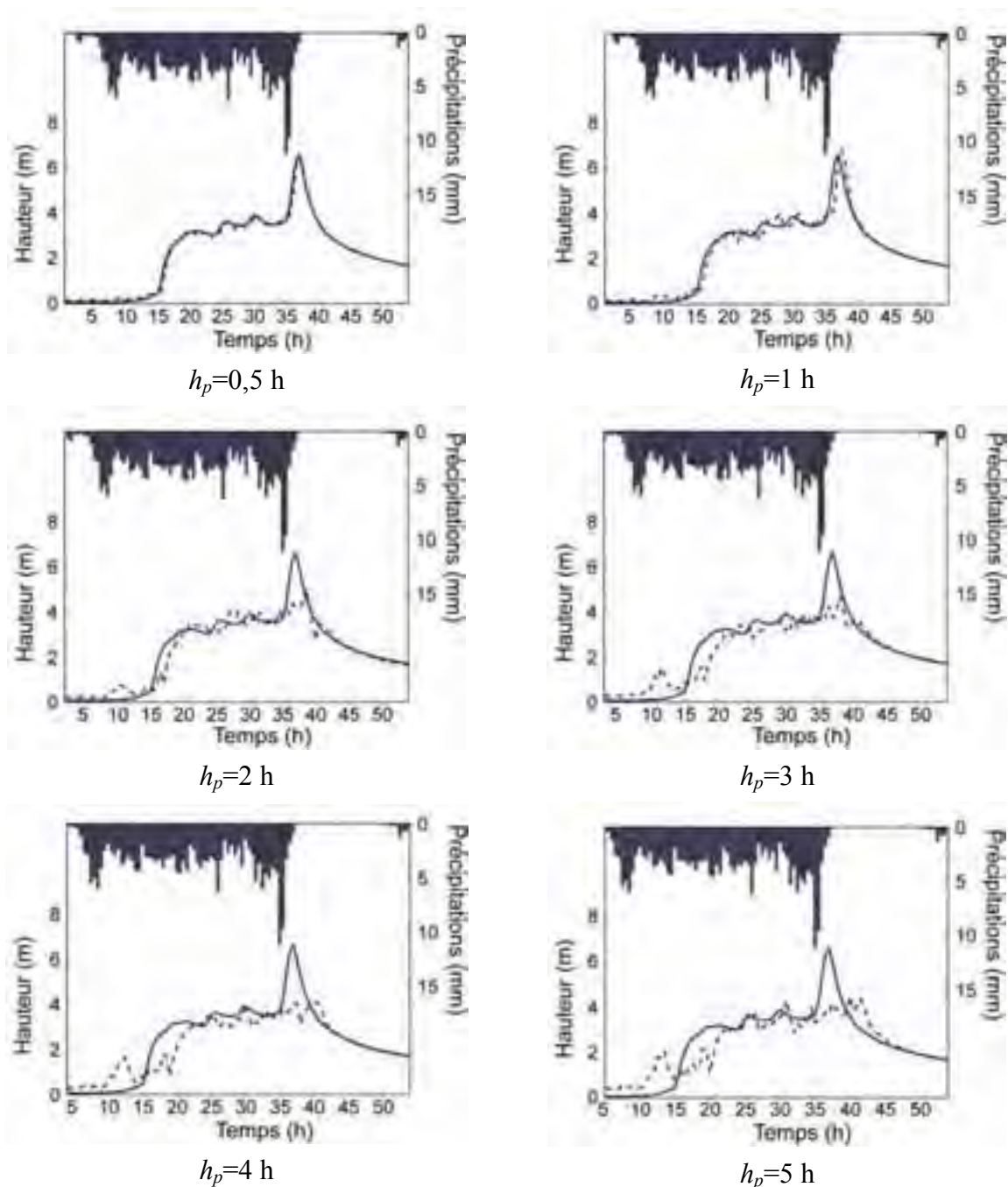
Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	2	5	3	3	3	2
Critères de persistance $C_p$	0,72	0,57	0,38	0,78	0,67	0,64	0,63
$R^2$ (Critère de Nash)	1,00	0,99	0,93	0,95	0,88	0,82	0,93
Critère de hauteur $C_h$	0,99	0,95	0,89	0,93	0,91	0,84	0,92

**Tableau 17 : Critères de qualité des prévisions sur l'événement 107**

### **Événement 23 d'octobre 2006.**

L'événement 23 d'octobre 2006 est un événement assez atypique avec un long plateau de crue moyenne et l'arrivée en fin d'événement d'un pic du type de celui de 2002. Heureusement le pic s'est arrêté à 6 m de hauteur environ. On pourra remarquer que le prédicteur a beaucoup de mal à prévoir cet événement à partir de l'horizon 2 h puisqu'il ne « voit » pas du tout de pic à partir de cet horizon. On peut expliquer simplement ce comportement : il n'y a en effet que 2 impulsions de pluies très intenses (de l'ordre de 10 mm en ½ h), qui se produisent moins de 2 h avant le pic ; le prédicteur n'a donc pas l'information des précipitations avant d'effectuer sa prévision. De plus, rien dans l'évolution des précipitations ou de la hauteur, qui étaient stabilisées depuis une dizaine d'heures, ne pouvait laisser envisager ce bref pic de précipitations.

Les critères de qualité sont paradoxalement bons à tous les horizons de prévision car les erreurs, même importantes, du pic, affectent peu les résultats moyens.



**Figure 36 : Limnigrammes de test sur l'événement 23 avec la validation croisée partielle et l'arrêt précoce sur l'événement 13**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$N_c$	2	2	5	3	3	3	2
Critères de persistance $C_p$	0,64	0,43	0,53	0,55	0,56	0,60	0,55
$R^2$ (Critère de Nash)	0,99	0,96	0,89	0,83	0,77	0,73	0,86
Critère de hauteur $C_h$	0,93	0,85	0,75	0,71	0,69	0,67	0,77

**Tableau 18 : Critères de qualité des prévisions sur l'événement 23**

#### *III.5.4.b Rapidité de convergence*

Les différentes méthodes de régularisation utilisées présentent également des différences en termes de vitesse de convergence. Rappelons que :

- l'apprentissage avec la validation croisée sans régularisation active est arrêté à 50 itérations,
- l'apprentissage avec l'arrêt précoce, dans toutes les configurations, est arrêté lorsque le critère d'arrêt atteint son minimum,
- l'apprentissage avec la seule modération des poids se poursuit durant 50 itérations mais en pratique atteint des incréments presque nuls à une certaine itération antérieure au maximum. C'est cette itération qui est reportée dans le tableau ci-dessous.

On constate de manière inattendue que la méthode la plus rapide n'est pas celle de l'arrêt précoce seul, mais celle qui combine l'arrêt précoce et la modération des poids. Cependant toutes les méthodes avec régularisation active convergent rapidement ; c'est, nous l'avons vu la condition qui évite la surparamétrisation.

Horizon de prévision $h_p$ (heures)		0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Nombres d'itérations avant convergence	sans régularisation active	50	50	50	50	50	50	50
	avec régularisation par arrêt précoce	7	7	7	7	7	7	7
	partielle avec arrêt précoce	9	6	5	11	20	6	10
	partielle avec modération des poids	11	6	4	4	4	4	6
	partielle avec arrêt précoce et modération des poids	4	5	3	3	5	5	4

**Tableau 19 : Rapidité de convergence.**

### **III.6 Conclusion**

Le troisième chapitre de notre mémoire a débuté sur le constat, issu de la bibliographie, que l'apprentissage statistique pouvait difficilement être appliqué aux phénomènes extrêmes en hydrologie. Les motifs sont en effet sérieux : les phénomènes extrêmes sont par définition rares et les moyens de mesures des phénomènes naturels sont entachés de bruits et d'erreur parfois très importants.

Pour surmonter ces difficultés, nous avons proposé d'utiliser des méthodes connues de régularisation. L'application de la validation croisée s'est révélée inefficace du fait du manque d'événements intenses dans la base de données : il a donc fallu concevoir et mettre en œuvre la validation croisée partielle, opérant sur les seuls événements intenses. L'arrêt précoce à partir de l'événement le plus facile à prévoir de toute la base comme événement d'arrêt s'est révélé très efficace. L'utilisation simultanée des deux méthodes (validation croisée et arrêt précoce) a abouti à la définition d'un prédicteur neuronal efficace qu'il n'a pas été possible d'améliorer avec la régularisation par modération des poids.

Une méthodologie rigoureuse a donc été définie, pas à pas, au cours de ce chapitre, et a été appliquée à la prévision de la crue exceptionnelle du 9 septembre 2002 à Anduze. Ainsi, des prévisions de bonne qualité sur un événement d'intensité de pointe 50% supérieure à tous les autres ont pu être proposées. En revanche nous regrettons de n'avoir pas eu le temps d'approfondir la comparaison entre la modération des poids et l'arrêt précoce sur cette application particulière qui conduit à des comportements de l'apprentissage très atypiques.

Finalement, afin de s'assurer que les différentes spécialisations opérées sur les événements intenses n'ont pas rendu le prédicteur inapte à la prévision d'événements peu intenses, la

méthodologie a été appliquée à un autre événement possédant plusieurs pics, et de faible intensité. Cette validation présente également de bons critères de qualité.

### **III.7 Références bibliographiques du chapitre III**

- [COULIBALY *et al.* 1999] P. COULIBALY, F. ANCTIL, B. BOBEE (1999). "Prévision hydrologique par réseaux de neurones : Etat de l'art". Can. J. Civil Engng, 26(3), pp. 293-304.
- [COULIBALY *et al.* 2000 a] P. COULIBALY, F. ANCTIL, P. RASMUSSEN, B. BOBEE (2000). « A recurrent Neural Networks approach using indices of low frequency climatic variability to forecast regional annual runoff. Hydrological Processes 14:2755-2777.
- [COULIBALY *et al.* 2000 b] P. COULIBALY, F. ANCTIL, B. BOBEE (2000). "Daily reservoir inflow forecasting using artificial neural networks with stopped training approach" Journal of Hydrology, Volume 230, Issues 3-4, 8 pp. 244-257.
- [DARTUS *et al.* 1993] D. DARTUS, L. DEDECKER, J.-M. COURIVAUD (1993). « Utilisation d'un réseau neuronal pour l'étude de la propagation d'une onde de crue dans un canal ». Journal of Hydraulic Research, vol. 31, n°2, pp. 161-170.
- [GAUME *et al.* 2003] E. GAUME, R. GOSSET (2003). « Over-Parametrisation, a major obstacle to the use of artificial neural networks in hydrology ? ». Hydrology and Earth System Sciences. 7, 693-206.
- [GOVINDARAJU 2000 a] R.S. GOVINDARAJU (2000). "Artificial neural networks in hydrology. I: Preliminary concepts". Journal of Hydrologic Engineering, 5(2), 115-123.
- [GOVINDARAJU 2000 b] R.S. GOVINDARAJU, (2000). "Artificial neural networks in hydrology. II: Hydrologic applications". Journal of Hydrologic Engineering, 5(2), 124-137.
- [JOHANNET *et al.* 1994] A. JOHANNET, A. MANGIN, D. D'HULST (1994). « Subterranean Water Infiltration Modelling by Neural Networks: Use of Water Source Flow ». In Proc. of ICANN; pp. 1033-1036. M. Marinaro and P.G. Morasso eds, Springer.
- [KARUNANITHI *et al.* 1994] N. KARUNANITHI, W. J. GRENNY, D. WHITLEY, K. BOVEE (1994). "Neural Networks for River Flow Prediction". J. Comp. in Civ. Engrg. 8, 201.
- [PIOTROWSKY *et al.* 2006] A. PIOTROWSKY, J.-J. NAPIORKOWSKI, P.-M. ROWINSKI, (2006). « Flash Flood Forecasting by Means of Neural Networks and Nearest Neighbour Approach - A Comparative Study. Nonlin. Processes Geophys. 13, 443-48.
- [SAHOO *et al.* 2006 a] G-B. SAHOO, C. RAY, (2006). « Flow forecasting for a Hawaii stream using rating curves and neural Networks ». Journal of Hydrology 317, 63-80.
- [SAHOO *et al.* 2006 b] G-B. SAHOO, C. RAY, E-H. DE CARLO, (2006). « Use of neural networks to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii ». Journal of Hydrology 327, 525-538.
- [SUDHEER *et al.* 2002] K. P. SUDHEER, A. K. RANGAN, S. M. SAHEB, (2002). « Modelling evaporation using an artificial neural network algorithm ». Hydrological Processes.16, 3189-3202.



[SCHMITZ *et al.* 2008] G.H. SCHMITZ, J. CULLMANN (2008). “PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments”. *Journal of Hydrology* 360, 1-14.

[ZEALAND *et al.* 1999] C M. ZEALAND, D. H. BURN, S. P. SIMONOVIC (1999): “Short term streamflow forecasting using artificial neural networks”. *Journal of Hydrology* 214, 32-48.

## Conclusion Générale

Les travaux réalisés durant cette thèse ont cherché à dépasser la contradiction apparente entre l'apprentissage statistique qui nécessite une base de données étendue, décrivant de manière représentative toutes les zones de l'espace de fonctionnement du processus que l'on cherche à modéliser, et le calage d'un modèle pour l'hydrologie devant réaliser avec suffisamment de fiabilité la prévision d'un événement extrême. Rappelons que l'événement de septembre 2002, s'il n'est pas unique dans l'histoire - le dernier événement comparable date de 1958 - est unique dans la base de données. En effet par rapport au second événement le plus intense de la base, l'événement de 2002 est 50 % plus intense.

L'autre difficulté à laquelle a dû faire face ce travail est l'intensité des bruits de mesure et des erreurs d'estimation de la pluie délivrée par ces orages cévenols cataclysmiques, extrêmement hétérogènes tant dans l'espace que dans le temps.

La présentation de l'apprentissage statistique nous a permis de mettre en évidence les méthodes connues qui permettent d'aborder ces deux difficultés. La validation croisée permet d'estimer l'erreur de généralisation sur l'ensemble de la base, et ainsi de réaliser la sélection du modèle ; les méthodes de régularisation permettent de résoudre le dilemme biais-variance, c'est-à-dire de réaliser l'apprentissage de la fonction qui sous-tend le processus observé, et d'ignorer les réalisations particulières du bruit intégrées dans les mesures.

Les deux méthodes de régularisation utilisées - la modération des poids et l'arrêt précoce - ont montré leur efficacité, mais il a fallu les adapter pour parvenir à estimer le comportement très intense d'un événement extrême. Pour cela, il a fallu éviter les phénomènes de spécialisation observés en fonction de l'ensemble d'arrêt, et introduire la validation croisée partielle, qui est en fait spécialisée sur les événements intenses. C'est grâce à l'établissement d'une typologie des événements, et grâce à une distribution judicieuse de ceux-ci dans les différents sous-ensembles d'apprentissage, de validation, d'arrêt et de test, que nous avons pu établir une procédure générale pour définir ces sous-ensembles et concevoir le modèle de prévision.

Les performances du modèle obtenu en prévision permettent d'envisager qu'une alerte fiable soit disponible sur Internet par l'intermédiaire du SCHAPI et de son site *vigicrues*, rendant ainsi une aide précieuse aux populations.

Contrairement à ce que l'on peut lire dans de nombreuses publications sur les réseaux de neurones appliqués à la prévision des crues, il n'est pas impossible de généraliser à un événement plus intense que ceux de l'ensemble d'apprentissage et il n'est pas non plus impossible d'éviter le surajustement dû au bruit dans les mesures ; nous l'avons démontré dans ce mémoire. C'est également pour cela que les perspectives ouvertes par ce travail sont particulièrement attrayantes autant pour intégrer de l'adaptativité dans le modèle que pour généraliser leurs applications à des bassins versants non jaugés.

Enfin une dernière piste de travail a été mise en évidence par le biais de ce travail ; si le modèle de prévision, conçu par apprentissage dirigé, que nous avons utilisé est le plus efficace, ceci peut signifier que les principales perturbations sur les mesures proviennent des entrées exogènes du prédicteur, c'est-à-dire des précipitations. Les travaux sur l'intégration des précipitations issues des radars atmosphériques sont donc tout à fait envisageables, ainsi que l'intégration directe des réflectivités issues des radars.

Le travail que nous avons présenté est une des bases du projet FLASH, soutenu par l'ANR dans son programme SYSCOMM.

## Annexe A : Méthode de Levenberg-Marquardt

La méthode d'estimation du gradient du second ordre selon l'algorithme de Levenberg-Marquardt utilise une approximation du hessien calculée à l'aide des produits des dérivées premières. La première étape est donc de calculer le gradient de la fonction de coût :

$$J(\mathbf{w}) = \sum_{k=1}^N (y^k - y(\mathbf{x}^k, \mathbf{w}))^2 = \sum_{k=1}^N J^k(\mathbf{w}), \text{ où } J^k(\mathbf{w}) \text{ est la fonction de coût calculée pour le seul exemple } k, N \text{ est le nombre d'exemples d'apprentissage, et } \mathbf{w} \text{ est le vecteur des paramètres.}$$

L'algorithme de la rétropropagation permet de calculer le gradient de la fonction de coût partielle  $J^k(\mathbf{w})$ , il se met en œuvre de la façon suivante :

- la propagation : les exemples sont présentés successivement au réseau ; pour chacun, le réseau calcule les potentiels, les sorties des neurones cachés et celles des neurones de sortie.
- le calcul des dérivées premières en commençant par le neurone de sortie s'exprime ainsi à la présentation de l'exemple  $k$  ( $s_i$  est la sortie du neurone  $i$  (neurone caché ou neurone de sortie) et  $v_i$  est son potentiel ;  $v_i^k$  est le potentiel du neurone  $i$  évalué à la présentation de l'exemple  $k$ ) :

$$\frac{\partial J^k(\mathbf{w})}{\partial w_{ij}} = r_i^k \cdot s_j^k$$

avec :

$$r_i^k = -2(y^k - y(\mathbf{x}^k, \mathbf{w})) \left( \frac{\partial s_i}{\partial v_i} \right)_{v_i=v_i^k} \text{ pour les paramètres liés au neurone de sortie } i.$$

$$\text{ou } r_i^k = \sum_h r_h^k w_{hi} \left( \frac{\partial s_i}{\partial v_i} \right)_{v_i=v_i^k} \text{ pour les paramètres d'entrée des neurones cachés, } h \text{ est}$$

l'indice des neurones sur lesquels agit le neurone  $i$  (dans le cas d'un réseau à une seule couche cachée,  $h$  représente donc les neurones de la couche de sortie).

Les dérivées premières sont calculées à l'issue de la présentation des  $N$  exemples d'apprentissage par l'addition des dérivées ainsi obtenues :

$$\frac{\partial J}{\partial w_{ij}} = \sum_{k=1}^N \frac{\partial J^k(\mathbf{w})}{\partial w_{ij}}.$$

S'inspirant du principe des méthodes de gradient du second ordre, la règle de Levenberg-Marquardt calcule les paramètres  $\mathbf{w}^{(i+1)}$  selon l'expression suivante :

$$\mathbf{w}^{i+1} = \mathbf{w}^i - \left[ \tilde{\mathbf{H}} \mathbf{w}^i + \lambda_{i+1} \mathbf{I} \right]^{-1} \nabla J(\mathbf{w}^i)$$

où :

- $\tilde{\mathbf{H}} \mathbf{w}^i$  est l'approximation du Hessien  $\mathbf{H}$  de la fonction de coût  $J(\mathbf{w}^{(i)})$  à la présentation  $i$  de l'ensemble complet d'apprentissage,  $\tilde{\mathbf{H}} \mathbf{w}^i$  est la matrice dont les termes sont :

$$\tilde{h}_{ij,lm} \mathbf{w}^i = \frac{\partial J(\mathbf{w}^{(i)})}{\partial w_{ij}} \frac{\partial J(\mathbf{w}^{(i)})}{\partial w_{lm}},$$

- $\mathbf{I}$  est la matrice identité,
- $\lambda_{\epsilon+1}$  est un paramètre qui assure le bon conditionnement de la matrice au démarrage de l'algorithme.

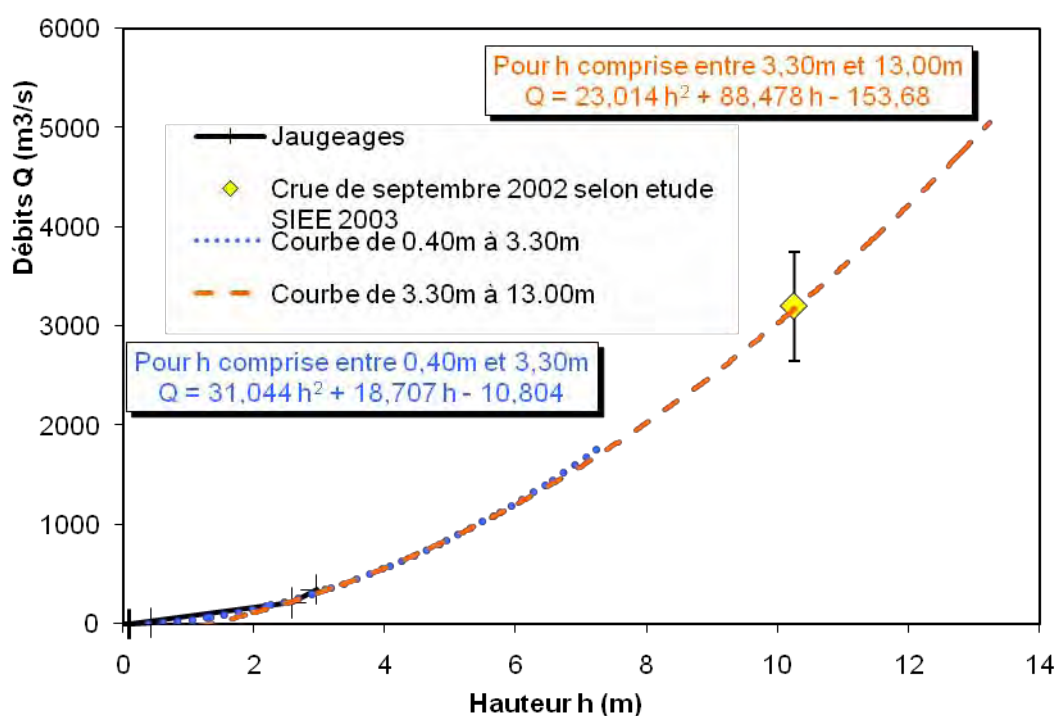
On pourra noter que  $\tilde{\mathbf{H}} \mathbf{w}^i$  ne fait intervenir que les dérivées premières de la fonction de coût dont on a vu précédemment qu'elles pouvaient être calculées aisément ; elle permet donc de diminuer la complexité de l'algorithme en évitant de calculer toutes les dérivées secondes (au nombre de  $p^2$ , s'il y a  $p$  paramètres).

Le paramètre  $\lambda_{\epsilon+1}$  représente, en quelques sortes, l'inverse du pas du gradient du premier ordre à l'itération  $(i+1)$  : au début de l'apprentissage, le paramètre  $\lambda_{\epsilon+1}$  est choisi grand afin

de rendre la matrice  $\left[ \tilde{\mathbf{H}} \mathbf{w}^i + \lambda_{i+1} \mathbf{I} \right]$  à diagonale dominante ; la méthode est alors équivalente à une méthode de premier ordre ; lorsque les paramètres se rapprochent du minimum de la fonction de coût,  $\lambda_{\epsilon+1}$  est progressivement diminué et les termes du second ordre peuvent prendre toute leur importance.

## Annexe B : Courbe de tarage

La courbe de tarage a pour objet de convertir les hauteurs d'eau mesurées à une station limnimétrique particulière en un débit à ce lieu précis. Pour la station de mesure d'Anduze, cette courbe est obtenue après plusieurs jaugeages réalisés pour plusieurs hauteurs. Malheureusement, compte tenu de l'ampleur des crues, on ne peut disposer de jaugeage pour les débits les plus importants, l'estimation de ce dernier pour de grandes hauteurs d'eau est donc très incertaine.



Courbe de tarage du Gardon d'Anduze à Anduze



## Annexe C : Origine des événements utilisés

Les événements numérotés de 1 à 19 sont les événements déjà connus et identifiés par ce numéro avant le début du projet BVNE. A ceux-ci ont été ajoutés les événements survenus depuis 2005 : les événements 22 à 25, ainsi que d'autres événements « anciens » identifiés dans les bases de données annuelles BVNE, moins intenses, numérotés à partir de la centaine. Les numéros 20 et 21, sont destinés aux événements de 2003 et 2005, non encore communiqués.

L'événement 19 est extrait du rapport SIEE, il s'agit des hauteurs corrigées (dénommées 19H dans le tableau ci-dessous) et du débit 19-Q1 de ce même tableau.

N°	date	Base de données	Fichier/feuille de calcul	Echantillonnage	Traitement effectué
1	21-24 septembre 1994	BVNE	annuel 1994	1h	Rééch. à ½ h
3	4-5 octobre 1995	BVNE	annuel 1995	1h	Rééch. à ½ h
4	13-14 octobre 1995	BVNE	annuel 1995	1h	Rééch. à ½ h
8	10-12 novembre 1996	BVNE	annuel 1996	1h	Rééch. à ½ h
13	28-29 septembre 2000	BVNE	annuel 2000	1h	Rééch. à ½ h
19bvne	9 septembre 2002	BVNE	Evènement seul	1h	NU
19ftp	9 septembre 2002	BVNE	FTP SCHAPI	5 min	Rééch. à ½ h
19-H	9 septembre 2002	SIEE	Anduze	5 min	Rééch. à ½ h
19-Q1	9 septembre 2002	SIEE	Anduze Charge Moy	5 min	Rééch. à ½ h
19-Q2	9 septembre 2002	SIEE	Anduze Charge Max	5 min	NU
19-Q3	9 septembre 2002	SIEE	Anduze Charge Min	5 min	NU
19-Q4	9 septembre 2002	SIEE	Anduze Moy	5 min	NU
19-Q5	9 septembre 2002	SIEE	Anduze Max	5 min	NU
19-Q6	9 septembre 2002	SIEE	Anduze Min	5 min	NU
22	24-25 septembre 2006	BVNE	FTP SCHAPI	¼ h	Rééch. à ½ h
23	19-20 octobre 2006	BVNE	FTP SCHAPI	¼ h	Rééch. à ½ h
24	17 novembre 2006	BVNE	FTP SCHAPI	¼ h	Rééch. à ½ h
25	20-23 novembre 2007	BVNE	FTP SCHAPI	¼ h	Rééch. à ½ h
107	5-7 novembre 1997	BVNE	annuel 1997	1h	Rééch. à ½ h
108	26-27 novembre 1997	BVNE	annuel 1997	1h	Rééch. à ½ h
109	18-19 décembre 1997	BVNE	annuel 1997	1h	Rééch. à ½ h
117	20-21 octobre 1999	BVNE	annuel 1999	1h	Rééch. à ½ h
120	12-14 novembre 2000	BVNE	annuel 2000	1h	Rééch. à ½ h

Liste, dénomination et provenance des données numériques utilisées (NU signifie Non Utilisé)



Les cumuls, durées et débits moyens sont calculés à partir des événements avec les instants de début et de fin reportés dans le tableau suivant (échantillonnage ½ h).

Evénements	Dates	
	Début	Fin
01	22/09/1994 20:08	24/09/1994 15:08
03	03/10/1995 23:08	06/10/1995 04:08
04	13/10/1995 14:09	15/10/1995 04:09
08	10/11/1996 10:00	13/11/1996 19:00
13	28/09/2000 09:00	30/09/2000 06:00
19	08/09/2002 13:00	10/09/2002 00:00
22	24/09/2006 04:00	25/09/2006 17:00
23	18/10/2006 11:00	20/10/2006 17:00
24	16/11/2006 22:45	18/11/2006 07:45
25	20/11/2007 22:15	23/11/2007 19:45
107	04/11/1997 03:00	07/11/1997 19:30
108	24/11/1997 19:30	28/11/1997 04:00
109	16/12/1997 10:30	20/12/1997 17:30
117	20/10/1999 14:00	21/10/1999 23:00
120	12/11/2000 02:00	15/11/2000 00:00

**Instants de début et de fin des événements**

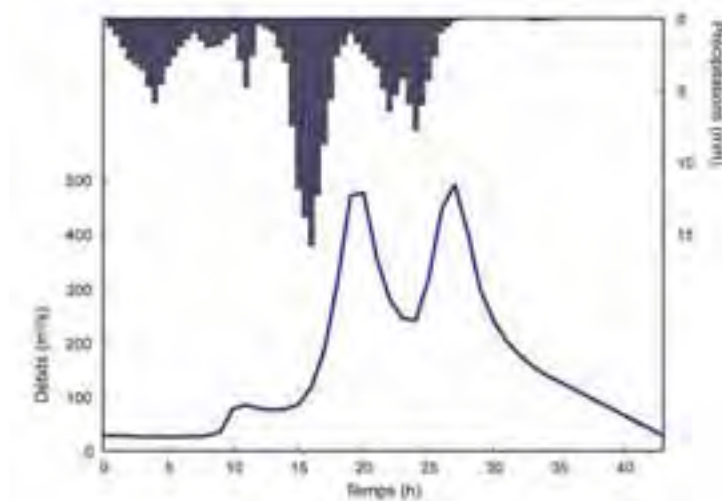
Lorsqu'un rééchantillonnage a été réalisé, ce dernier est mené à bien de la manière suivante.

- Lorsque la période d'échantillonnage augmente (par exemple de 5 min à 1/2h) :
  - On affecte au temps t la valeur de la moyenne des n valeurs précédentes pour les débits et les hauteurs, afin de rendre le filtre « passe bas » causal en vue d'une utilisation en temps réel.
  - On affecte le cumul des pluies des n valeurs précédentes pour les pluviomètres.
- Lorsque la période d'échantillonnage diminue (par exemple de 1h à ½ h) :
  - On ajoute une valeur interpolée des débits ou hauteurs.
  - On ajoute une valeur des cumuls divisée par le facteur de diminution de la période pour les cumuls de pluies.

## Annexe D : Présentation des événements de la base de données

### Événement n°1 du 21 au 24 septembre 1994

Cet événement est considéré comme non intense, court et comportant plusieurs pics.

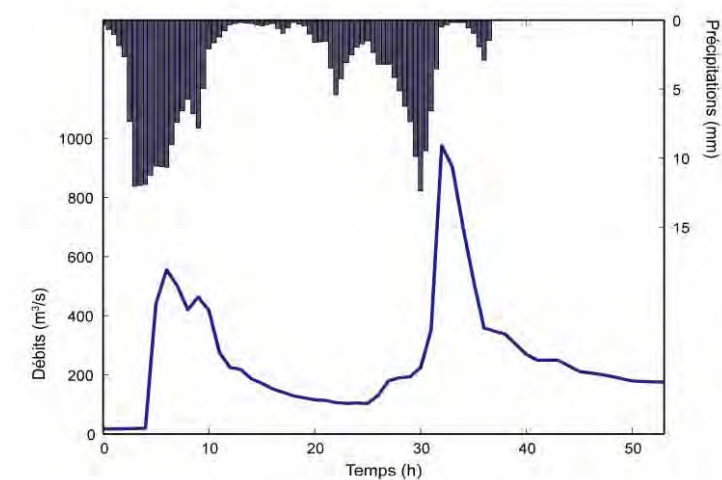


**Précipitations et limnigramme de l'événement 01**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°3 du 4-5 octobre 1995

Cet événement est considéré comme intense, long et possédant plusieurs pics. Cet événement fut la cause d'un décès.

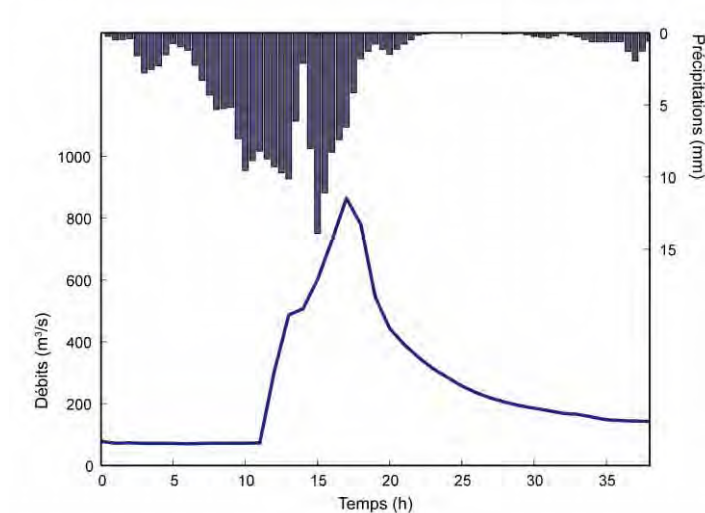


**Précipitations et limnigramme de l'événement 03**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

#### Evénement n°4 du 13 au 14 octobre 1995

Cet événement est considéré comme intense, long et possédant un seul pic.

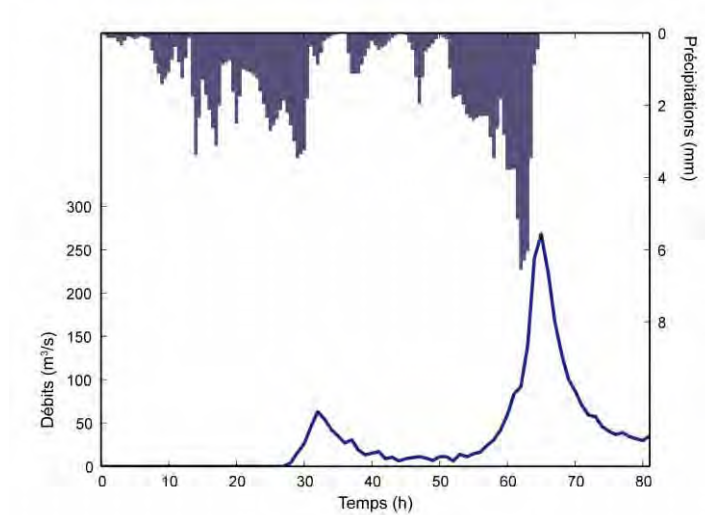


**Précipitations et limnigramme de l'événement 04**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

#### Evénement n°08 du 10 au 12 novembre 1996

Cet événement est considéré comme non intense, long et possédant plusieurs pics.

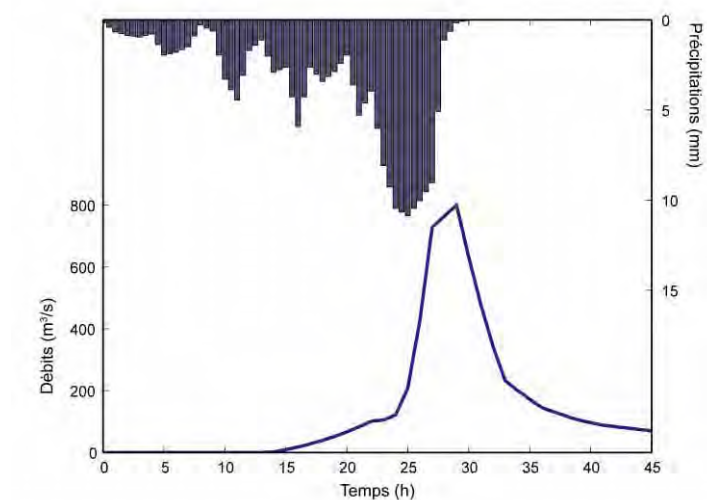


**Précipitations et limnigramme de l'événement 08**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°13 du 28 au 29 septembre 2000

Cet événement est considéré comme intense, court et possédant un seul pic.

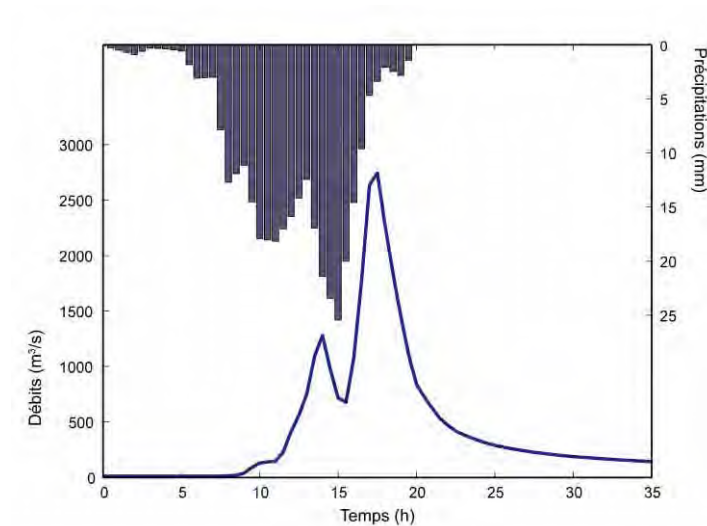


#### Précipitations et limnigramme de l'événement 13

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°19 du 8 au 9 septembre 2002

Cet événement est considéré comme intense, court et possédant un seul pic.

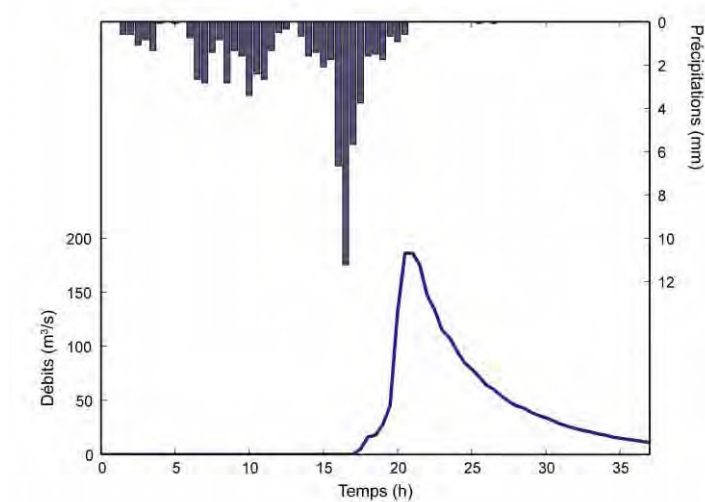


#### Précipitations et limnigramme de l'événement 19

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Evénement n°22 du 24 au 25 septembre 2006

Cet événement est considéré comme non intense, court et possédant un seul pic.

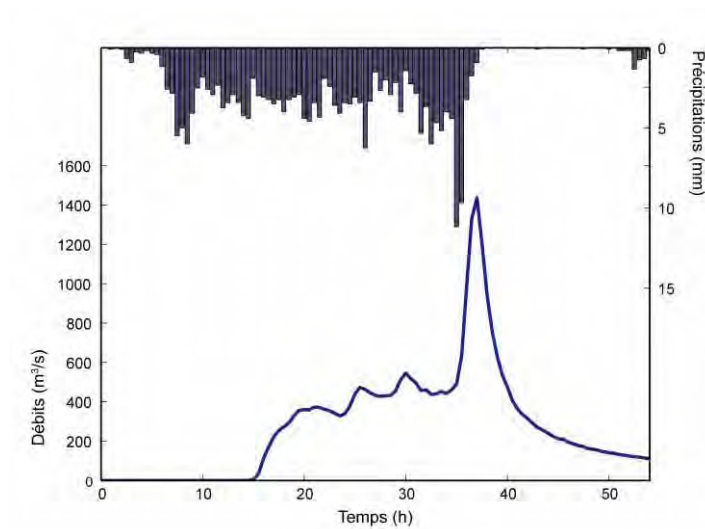


**Précipitations et limnigramme de l'événement 22**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Evénement n°23 du 19 au 20 octobre 2006

Cet événement est considéré comme intense long et possédant un seul pic.

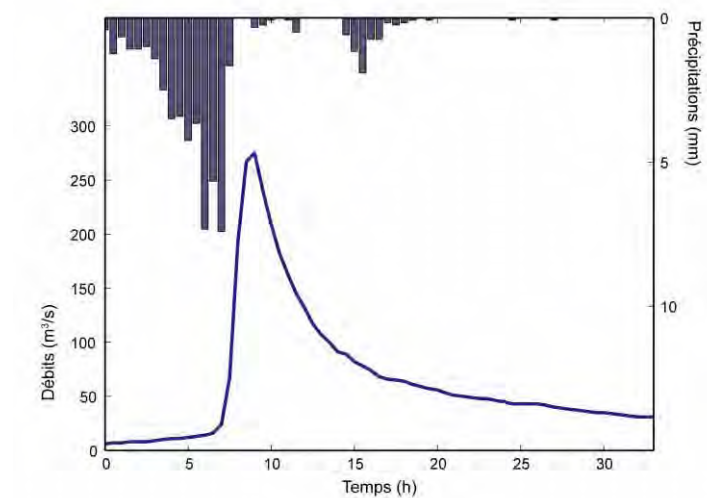


**Précipitations et limnigramme de l'événement 23**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°24 du 17 novembre 2006

Cet événement est considéré comme non intense, court et possédant un seul pic.

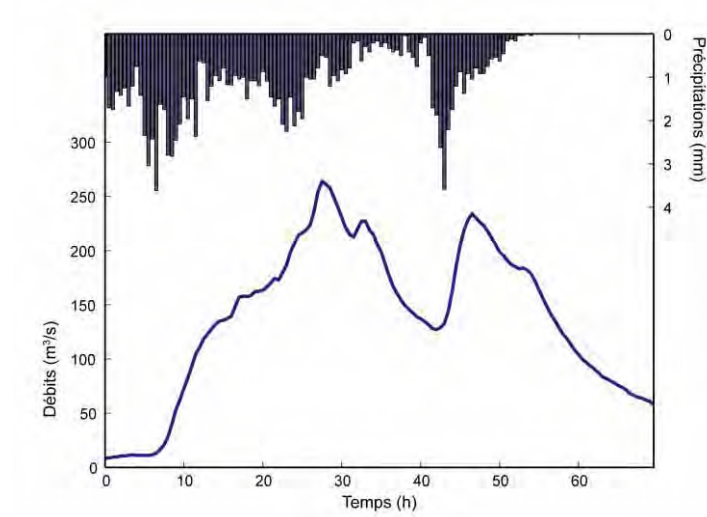


#### Précipitations et limnigramme de l'événement 24

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°25 du 20 au 23 novembre 2007

Cet événement est considéré comme non intense, long et possédant plusieurs pics.

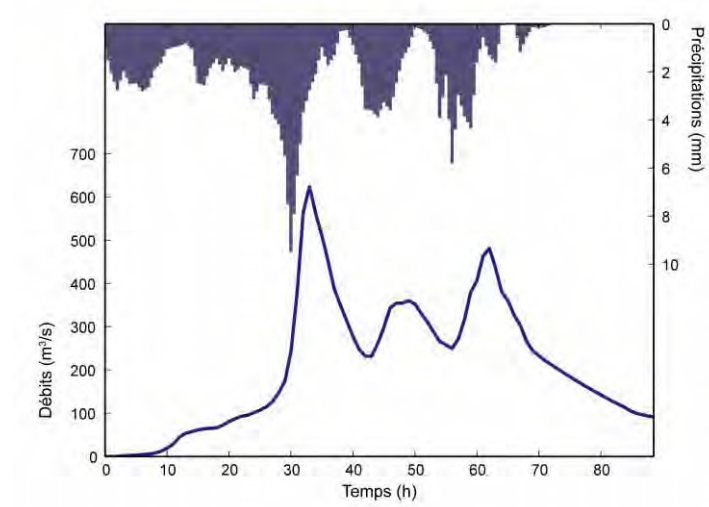


#### Précipitations et limnigramme de l'événement 25

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Evénement n°107 du 5 au 7 novembre 1997

Cet événement est considéré comme non intense, long et possédant plusieurs pics.

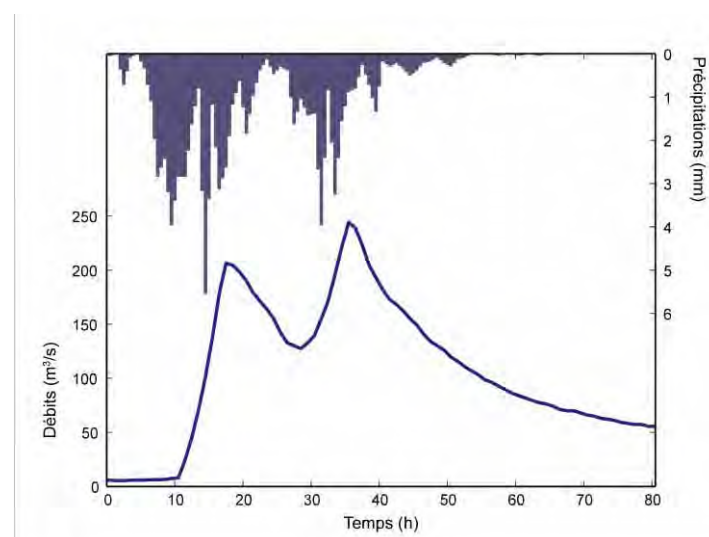


**Précipitations et limnigramme de l'événement 107**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Evénement n°108 du 26 au 27 novembre 1997

Cet événement est considéré comme non intense, long et possédant plusieurs pics.



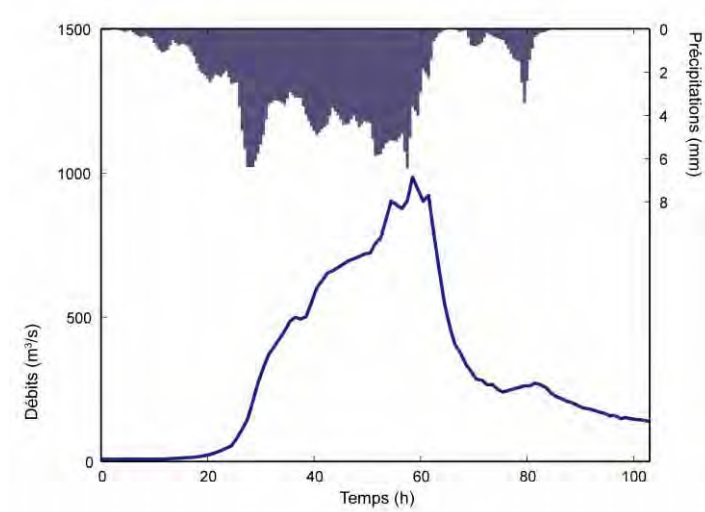
**Précipitations et limnigramme de l'événement 108**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.



### Événement n°109 du 18 au 19 décembre 1997

Cet événement est considéré comme intense long et comportant un seul pic.

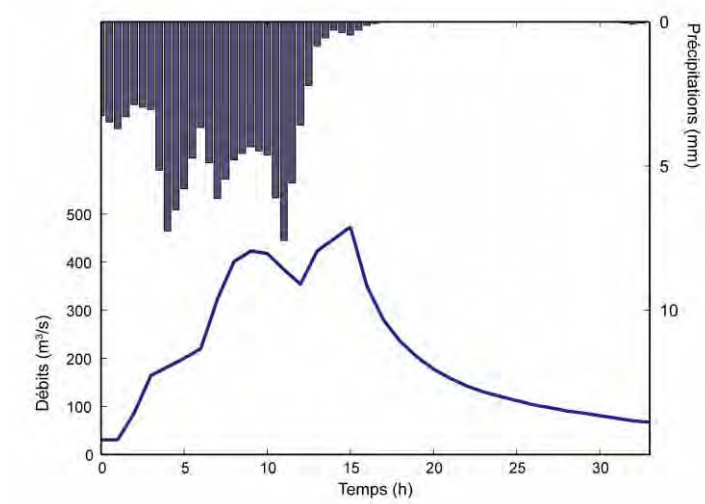


**Précipitations et limnigramme de l'événement 109**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

### Événement n°117 du 20 au 21 octobre 1999

Cet événement est considéré comme non intense court et comportant un seul pic.



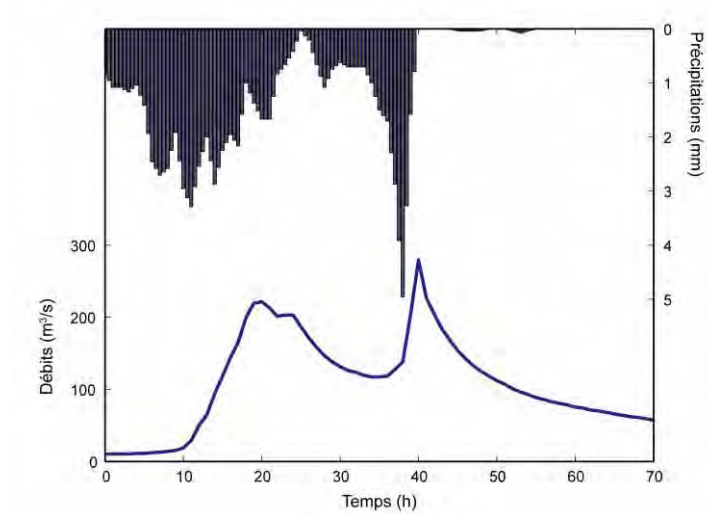
**Précipitations et limnigramme de l'événement 117**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.



### Événement n°120 du 12 au 14 novembre 2000

Cet événement est considéré comme non intense, long et possédant plusieurs pics.



**Précipitations et limnigramme de l'événement 120**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée.

## Annexe E : La crue exceptionnelle du 9 septembre 2002

Il est difficile de se représenter l'ampleur de la catastrophe de l'événement du 9 septembre 2002 dans tous les départements du Gard, de l'Hérault et de l'Ardèche, même si le phénomène a été le plus violent aux alentours d'Anduze. Cet événement a causé la mort de 23 personnes dans le Gard, 395 communes ont été déclarées en état de catastrophes naturelles dont 299 dans le Gard. 3000 entreprises, des exploitations agricoles et 7200 résidences ont été inondées. Les dégâts s'élèvent à 1,2 milliards d'euros. Les images ci-dessous permettent de mieux se figurer le pic de la crue. Nous remercions ici l'Office du Tourisme d'Anduze qui a bien voulu nous communiquer ces illustrations.



Vue depuis l'amont du pont vieux d'Anduze, le pont qui enjambe le Gardon est presque submergé, il est environ 8 h du matin et le débit avoisine  $2500 \text{ m}^3/\text{s}$ , le maximum est estimé à  $2800 \text{ m}^3/\text{s}$  à 8h30 du matin.



Cette fois ci vu depuis l'aval. On remarque les cascades qui dévalent la colline en arrière plan.



Photo cadrée légèrement à gauche de la précédente, le quai qui amène vers la ville, inondée par le ruissellement dévalant les collines.



La rue de l'écluse vers l'aval d'Anduze, on note en arrière plan de la photo la vague qui se situe dans le cours principal du Gardon, la digue ici ne protège plus la ville. Au fond on distingue les structures du stade d'Anduze qui se situe de l'autre côté du Gardon.



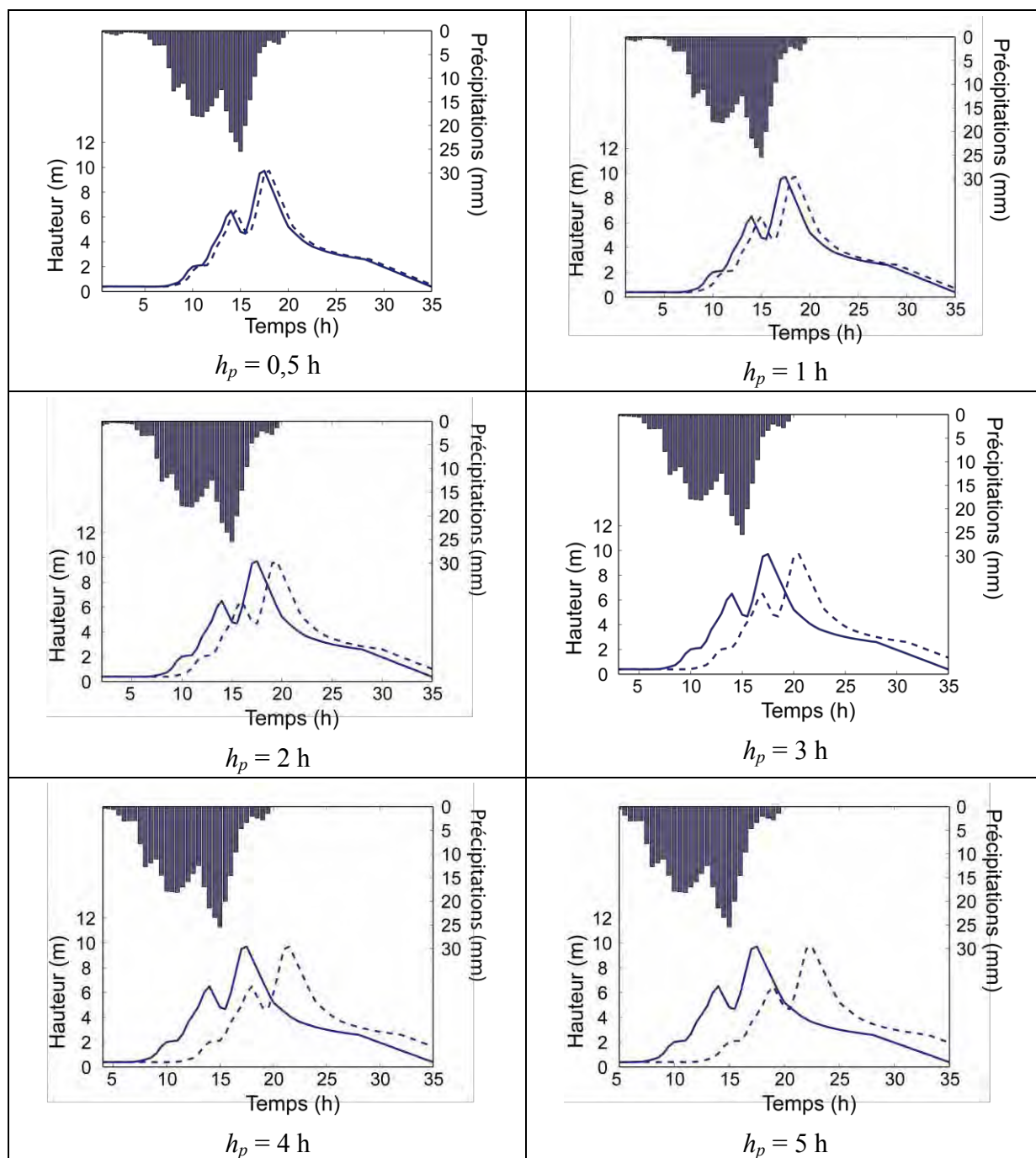
## Annexe F : Prévisions « naïves » de la crue de septembre 2002

Sont présentés ci-dessous les limnigrammes correspondant à la prévision naïve de l'événement de 2002. Les critères de performances  $R^2$  et critère de hauteur sont reportés dans le tableau ci-dessous. On peut noter que pour les faibles horizons de prévision, le coefficient de détermination est excellent, même si la prévision naïve n'apporte aucune information sur la hauteur d'eau à venir. Par définition, le critère de persistance de la prévision naïve est nul.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
$R^2$ (Critère de Nash)	0,95	0,84	0,57	0,34	0,07	-0,33	0,41
Critère de hauteur $C_h$	0,86	0,77	0,61	0,55	0,47	0,31	0,60

Critères de qualité des prévisions « naïves » de l'événement n°19





**Limnigrammes des prévisions « naïves » de l'événement n°19**

Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur correspondant à la prévision naïve.

## Annexe G : Meilleures prévisions obtenues

Cette annexe présente les prévisions obtenues lorsque la sélection du modèle est réalisée avec la validation croisée partielle et l'événement 13 de septembre 2000 en arrêt précoce. Le modèle a été sélectionné lorsque l'événement 19 était en test. Cependant, afin d'obtenir des prévisions en aveugle pour chacun des autres événements, le modèle précédemment sélectionné a été conservé et l'apprentissage a été réalisé à nouveau, chaque événement évalué étant en test et l'événement de 2002 en apprentissage. Pour une plus grande facilité de lecture, les résultats des événements 04, 23, 19 et 107 présentés dans le corps du chapitre III ont été repris dans cette annexe.

Le tableau ci-dessous présente les résultats synthétiques, on y reporte la catégorie à laquelle appartient l'événement (au sens de la classification présentée au chapitre II), puis si cet événement est mieux prévu, par rapport à la prévision naïve, en moyenne ou pour chacun des horizons de prévision.

N°	date	Intense	long	Mono pic	Classe	Prévision meilleure en moyenne que la prévision naïve	Prévision meilleure que la prévision naïve pour chaque $h_p$
<b>1</b>	21-24 septembre 1994	non	non	non	H	oui	oui
<b>3</b>	4-5 octobre 1995	oui	oui	non	B	oui	oui
<b>4</b>	13-14 octobre 1995	oui	oui	oui	A	oui	oui
<b>8</b>	10-12 novembre 1996	non	oui	non	F	oui	oui
<b>13</b>	28-29 septembre 2000	oui	non	oui	C	Ensemble d'arrêt	Ensemble d'arrêt
<b>19</b>	9 septembre 2002	oui	non	oui	C	oui	oui
<b>22</b>	24-25 septembre 2006	non	non	oui	G	oui	non (1/2h et 1 h)
<b>23</b>	19-20 octobre 2006	oui	oui	oui	A	oui	oui
<b>24</b>	17 novembre 2006	non	non	oui	G	oui	oui
<b>25</b>	20-23 novembre 2007	non	oui	non	F	oui	non (2 h)
<b>107</b>	5-7 novembre 1997	non	oui	non	F	oui	oui
<b>108</b>	26-27 novembre 1997	non	oui	non	F	oui	oui
<b>109</b>	18-19 décembre 1997	oui	oui	oui	A	oui	non (2 h et 3 h)



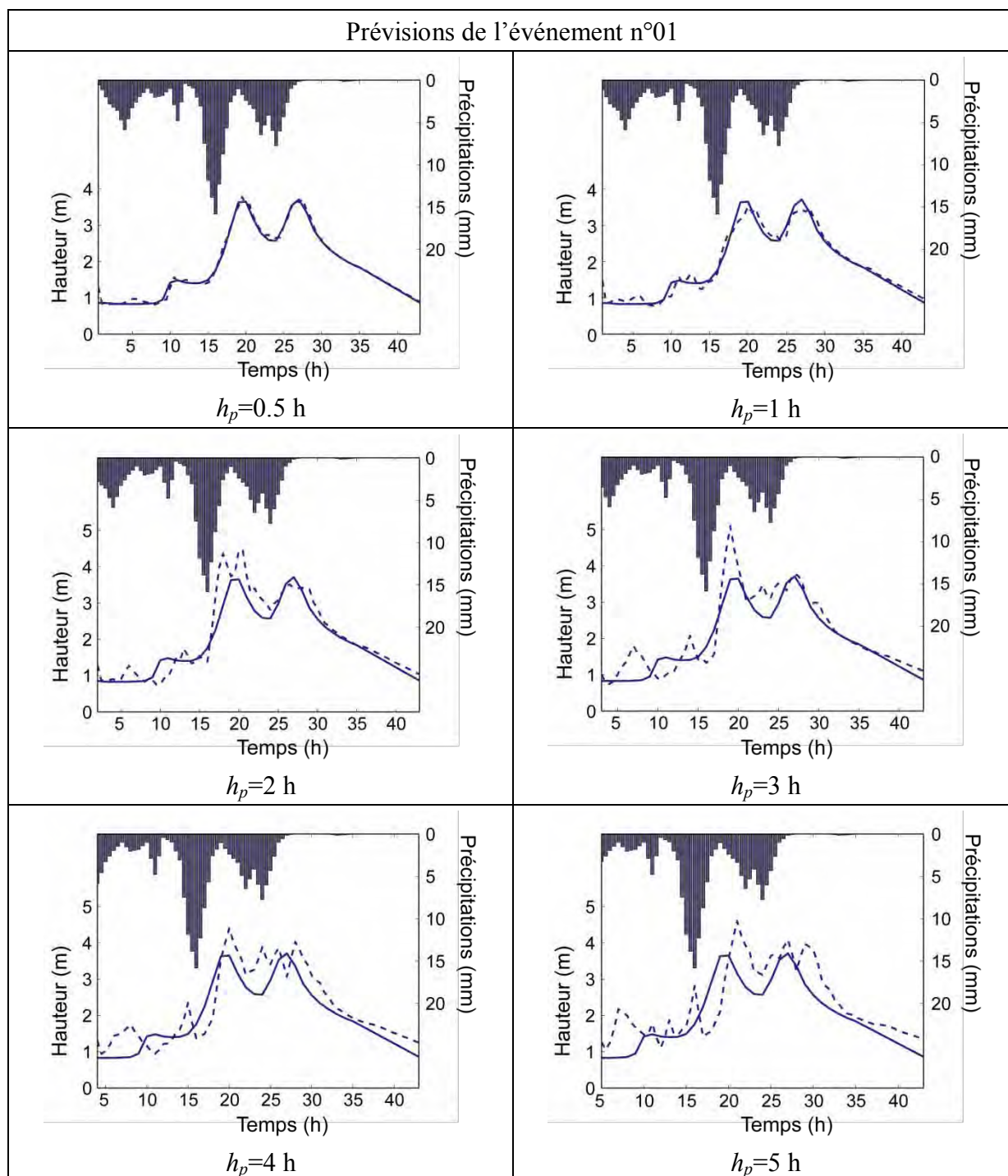
<b>117</b>	20-21 octobre 1999	non	non	oui	G	oui	oui
<b>120</b>	12-14 novembre 2000	non	oui	non	F	oui	oui

---

On peut remarquer que, en moyenne sur les horizons de prévisions, tous les événements sont mieux prévus qu'avec la prévision naïve. Néanmoins pour trois événements (25, 22, 109), certains horizons de prévision posent encore problème. On pourra noter que ces événements n'appartiennent pas à une classe particulière ; ils peuvent être intenses autant que non intenses. La validation croisée partielle ne semble donc pas avoir induit de spécialisation par rapport aux seuls événements intenses, mais en revanche a servi à mieux les prévoir.

Ci-dessous sont présentées les prévisions et critères de qualité pour tous les événements de la base de donnée hormis l'événement 13 qui a servi d'ensemble d'arrêt.

## Événement 01

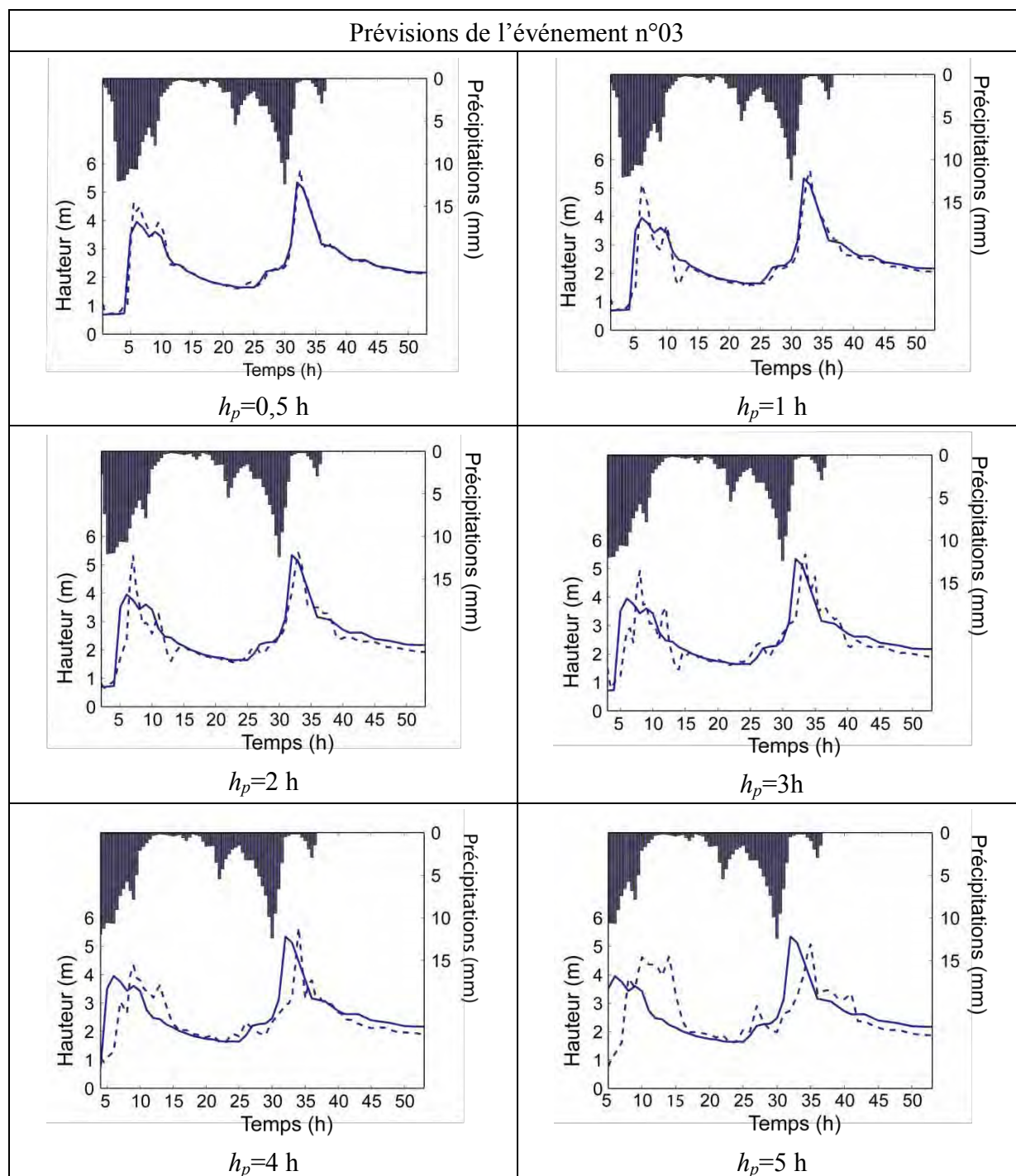


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,60	0,61	0,40	0,61	0,59	0,39	0,53
$R^2$ (Critère de Nash)	0,99	0,97	0,83	0,79	0,69	0,42	0,78
Critère de hauteur $C_h$	0,97	0,94	0,91	0,90	0,90	0,84	0,91
$R^2$ (prévision naïve)	0,98	0,92	0,71	0,46	0,23	0,05	0,56
$C_h$ (prévision naïve)	0,95	0,89	0,77	0,67	0,62	0,63	0,76

Critères de qualité de la prévision

## Evénement 03

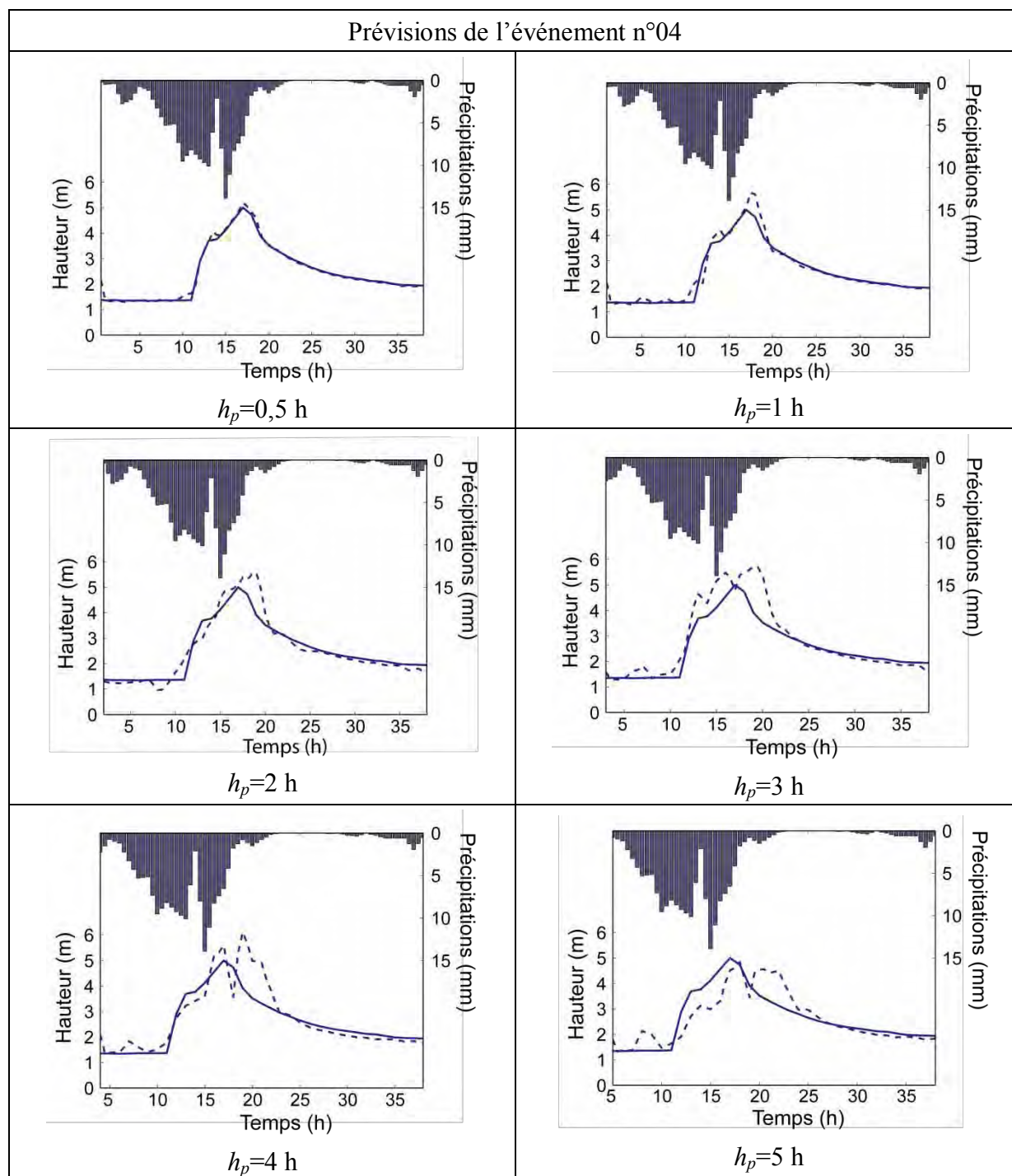


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,43	0,47	0,69	0,71	0,64	0,48	0,57
$R^2$ (Critère de Nash)	0,95	0,86	0,75	0,60	0,26	-0,37	0,51
Critère de hauteur $C_h$	0,86	0,81	0,75	0,70	0,55	0,47	0,69
$R^2$ (prévision naïve)	0,92	0,73	0,20	-0,40	-1,03	-1,61	-0,20
$C_h$ (prévision naïve)	0,81	0,67	0,46	0,40	0,39	0,33	0,51

Critères de qualité de la prévision

## Evénement 04



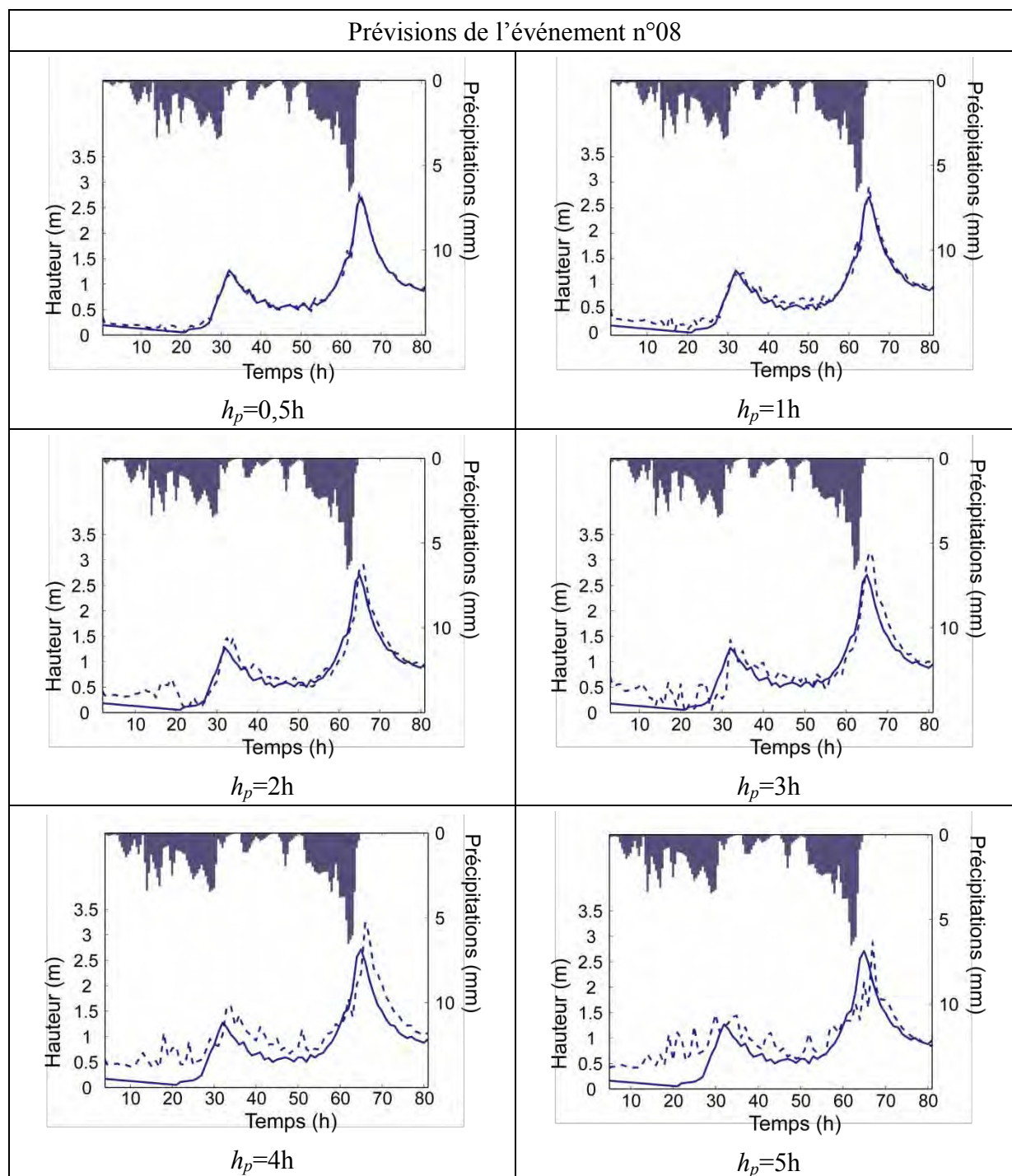
Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,48	0,44	0,60	0,51	0,66	0,78	0,58
$R^2$ (Critère de Nash)	0,99	0,94	0,87	0,70	0,68	0,69	0,81
Critère de hauteur $C_h$	0,98	0,96	0,94	0,88	0,92	0,83	0,92
$R^2$ (prévision naïve)	0,97	0,90	0,68	0,39	0,04	-0,39	0,43
$C_h$ (prévision naïve)	0,96	0,92	0,85	0,78	0,66	0,46	0,77

Critères de qualité de la prévision



## Événement 08



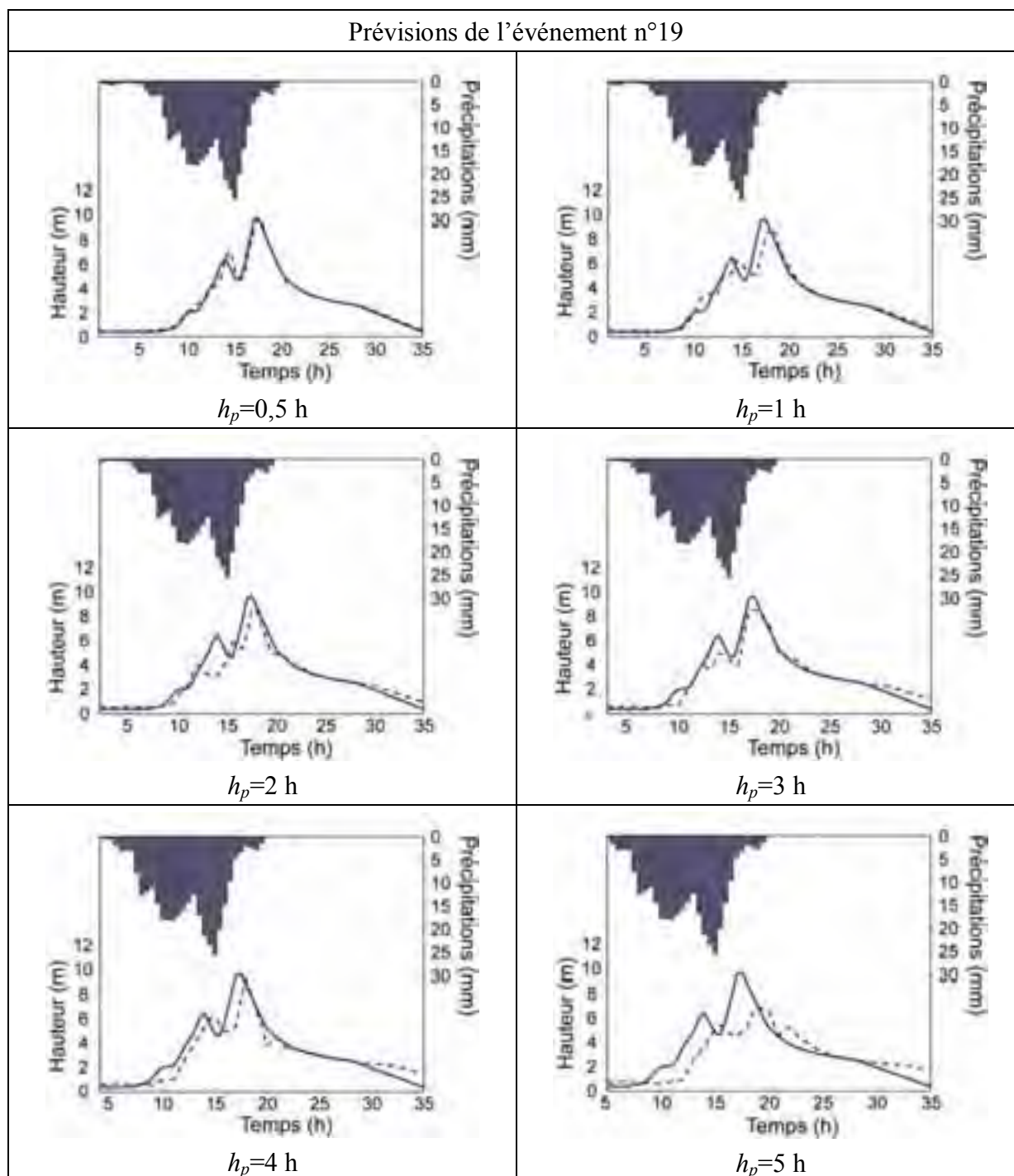
Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.



Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,30	0,14	0,25	0,41	0,14	0,30	0,26
$R^2$ (Critère de Nash)	0,99	0,96	0,88	0,82	0,60	0,55	0,80
Critère de hauteur $C_h$	0,97	0,96	0,89	0,80	0,83	0,80	0,88
$R^2$ (prévision naïve)	0,99	0,95	0,85	0,70	0,54	0,36	0,73
$C_h$ (prévision naïve)	0,92	0,85	0,68	0,54	0,42	0,32	0,62

Critères de qualité de la prévision

## Evénement 19

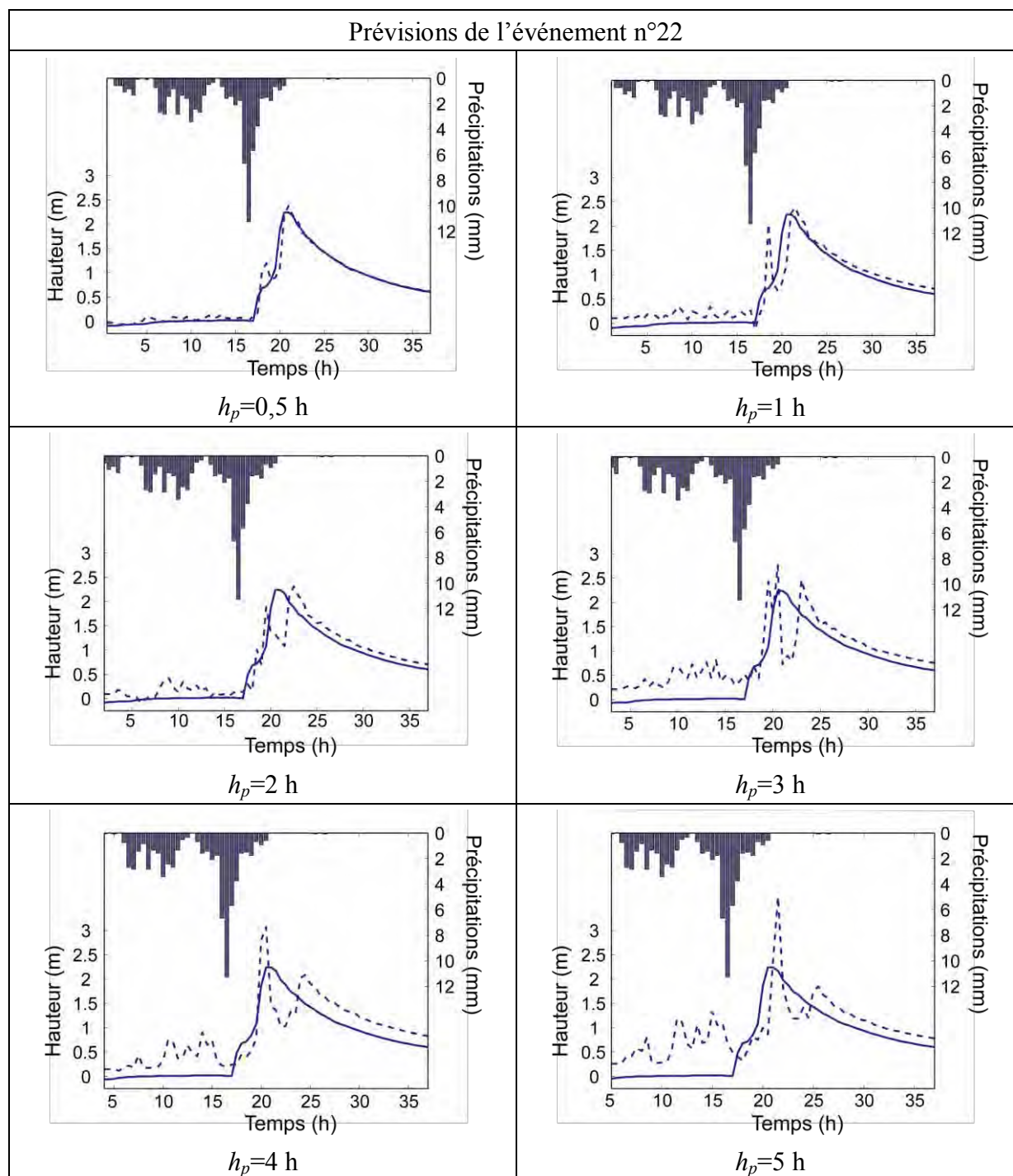


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,60	0,54	0,71	0,90	0,84	0,70	0,72
$R^2$ (Critère de Nash)	0,98	0,93	0,87	0,94	0,85	0,60	0,86
Critère de hauteur $C_h$	0,90	0,84	0,73	0,82	0,79	0,60	0,78
$R^2$ (prévision naïve)	0,95	0,84	0,57	0,34	0,07	-0,33	0,41
$C_h$ (prévision naïve)	0,86	0,77	0,61	0,55	0,47	0,31	0,60

Critères de qualité de la prévision

## Événement 22

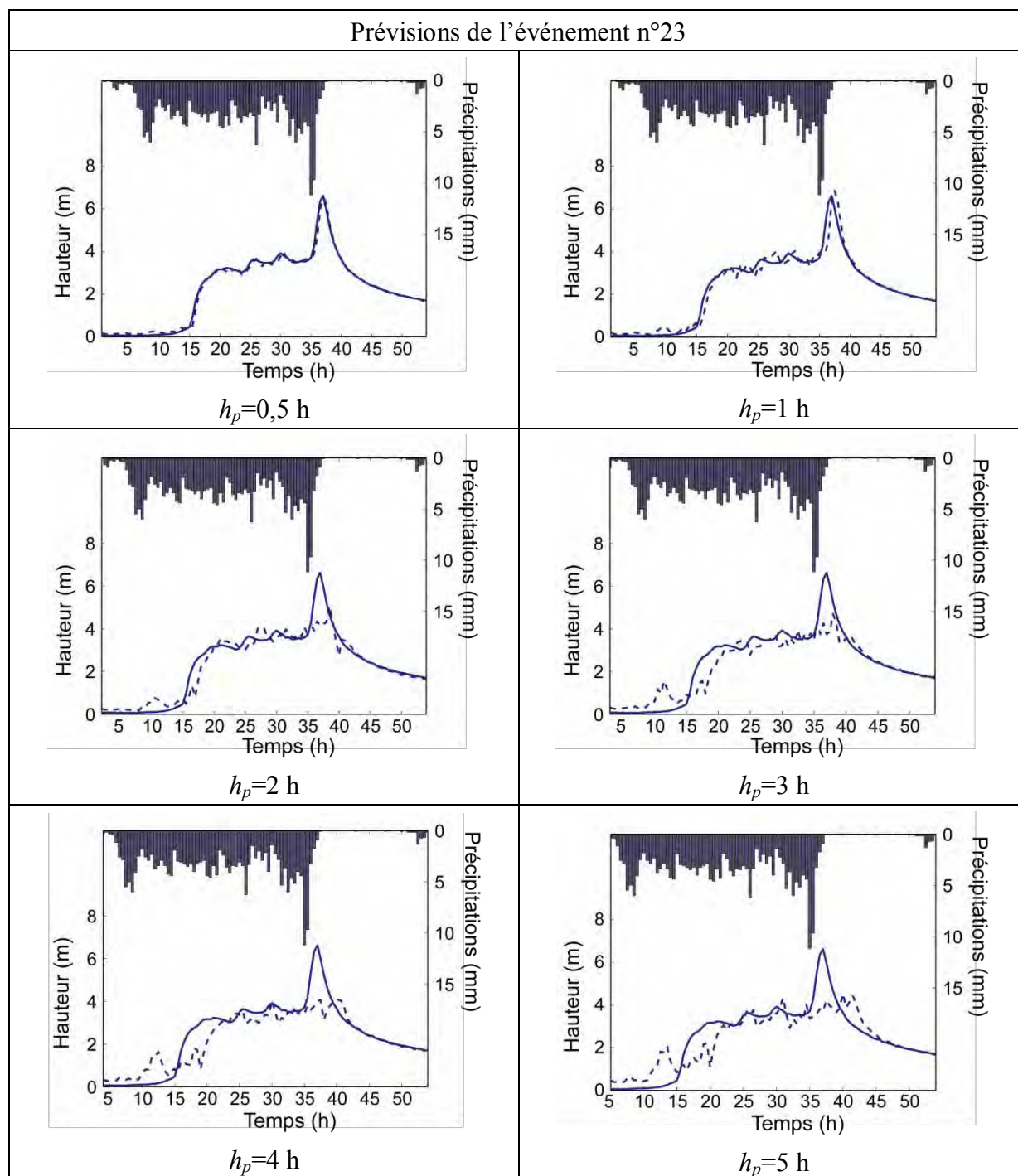


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	-0,06	-0,51	0,41	0,19	0,55	0,45	0,17
$R^2$ (Critère de Nash)	0,96	0,83	0,81	0,51	0,58	0,32	0,67
Critère de hauteur $C_h$	0,86	0,71	0,65	0,65	0,68	0,71	0,71
$R^2$ (prévision naïve)	0,96	0,89	0,67	0,40	0,07	-0,23	0,46
$C_h$ (prévision naïve)	0,81	0,65	0,30	0,11	0,01	0,01	0,32

Critères de qualité de la prévision

## Événement 23

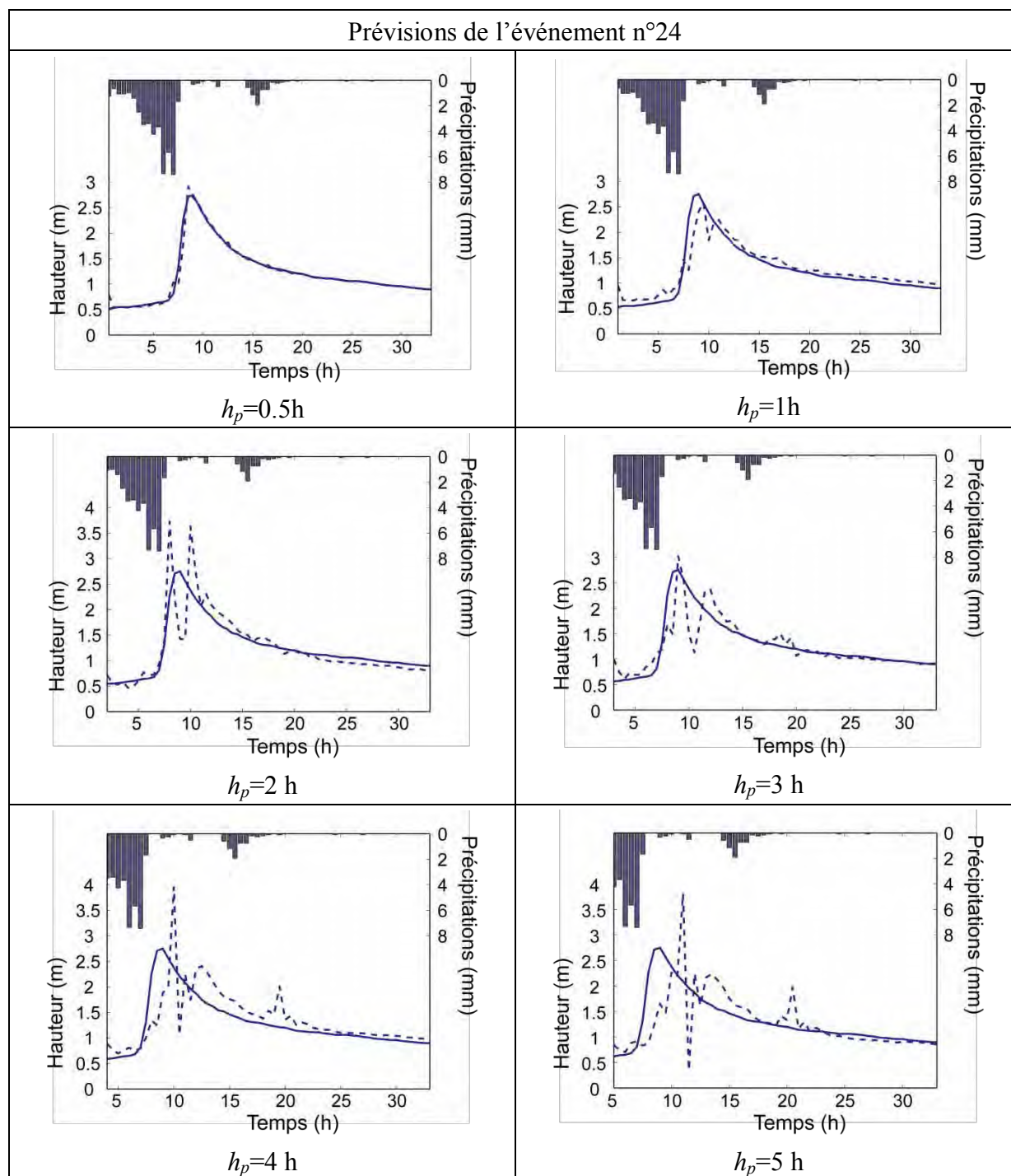


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,64	0,43	0,53	0,55	0,56	0,60	0,55
$R^2$ (Critère de Nash)	0,99	0,96	0,89	0,83	0,77	0,73	0,86
Critère de hauteur $C_h$	0,93	0,85	0,75	0,71	0,69	0,67	0,77
$R^2$ (prévision naïve)	0,98	0,92	0,77	0,62	0,47	0,31	0,68
$C_h$ (prévision naïve)	0,88	0,79	0,67	0,63	0,64	0,66	0,71

Critères de qualité de la prévision

## Événement 24



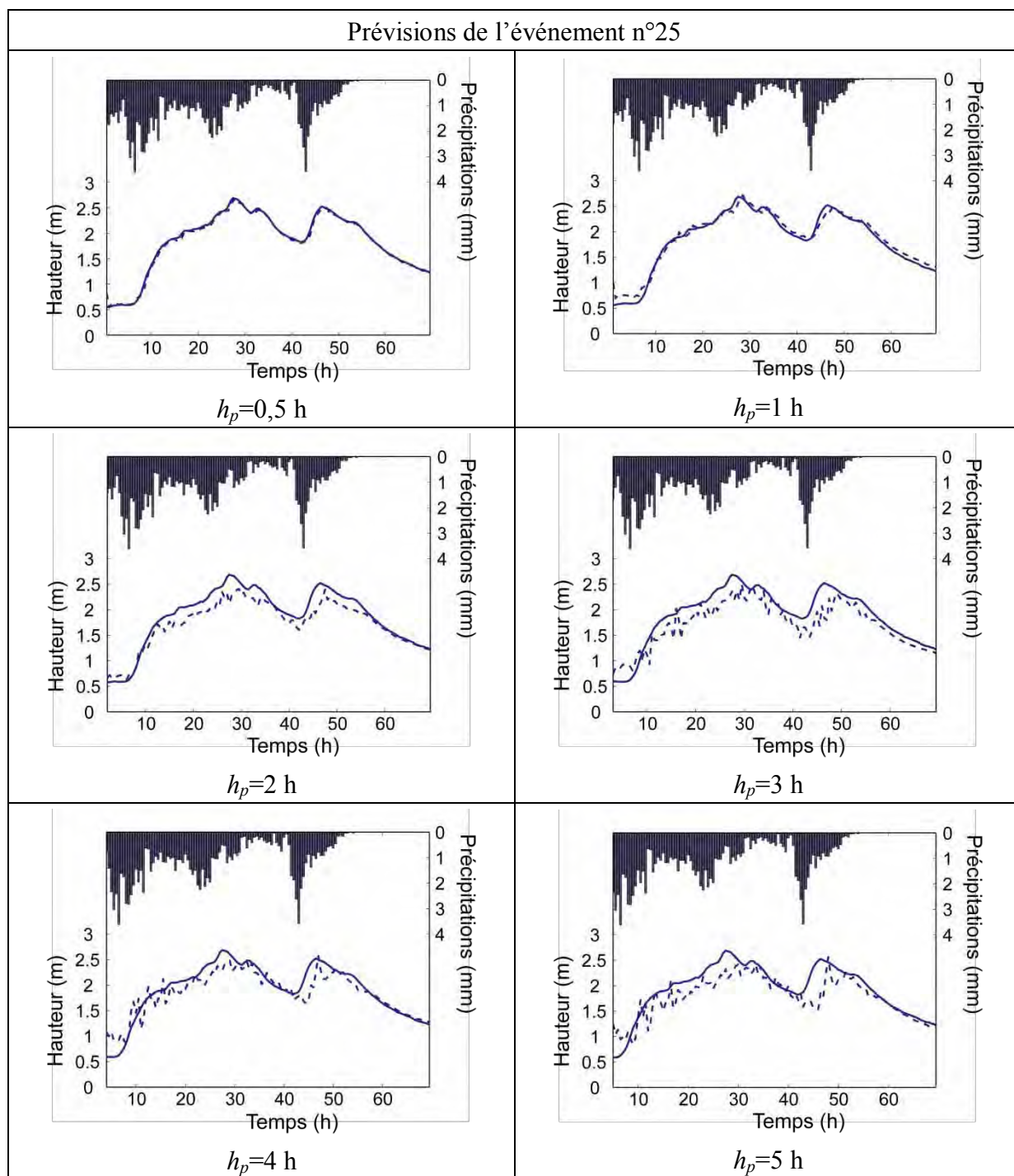
Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.



Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,65	0,49	0,46	0,77	0,63	0,58	0,60
$R^2$ (Critère de Nash)	0,97	0,85	0,55	0,69	0,29	0,00	0,56
Critère de hauteur $C_h$	0,88	0,81	0,69	0,82	0,65	0,53	0,73
$R^2$ (prévision naïve)	0,92	0,71	0,16	-0,38	-0,90	-1,40	-0,15
$C_h$ (prévision naïve)	0,79	0,63	0,37	0,29	0,28	0,26	0,44

Critères de qualité de la prévision

## Evénement 25

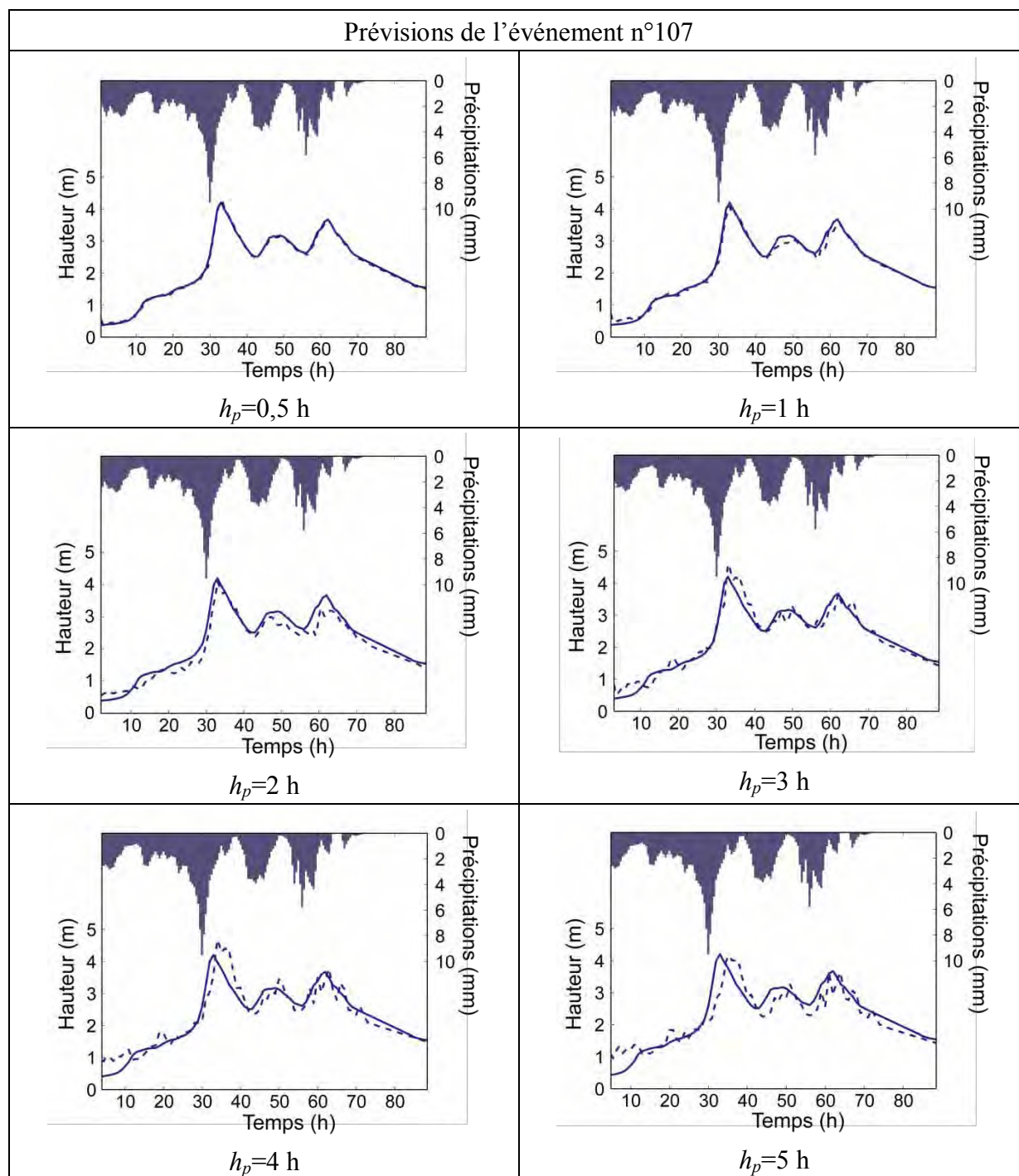


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,26	0,07	-0,14	-0,03	0,62	0,50	0,21
$R^2$ (Critère de Nash)	0,99	0,97	0,88	0,76	0,84	0,68	0,85
Critère de hauteur $C_h$	0,98	0,95	0,86	0,81	0,87	0,76	0,87
$R^2$ (prévision naïve)	0,99	0,97	0,89	0,76	0,58	0,34	0,76
$C_h$ (prévision naïve)	0,98	0,95	0,89	0,84	0,81	0,80	0,88

Critères de qualité de la prévision

## Evénement 107

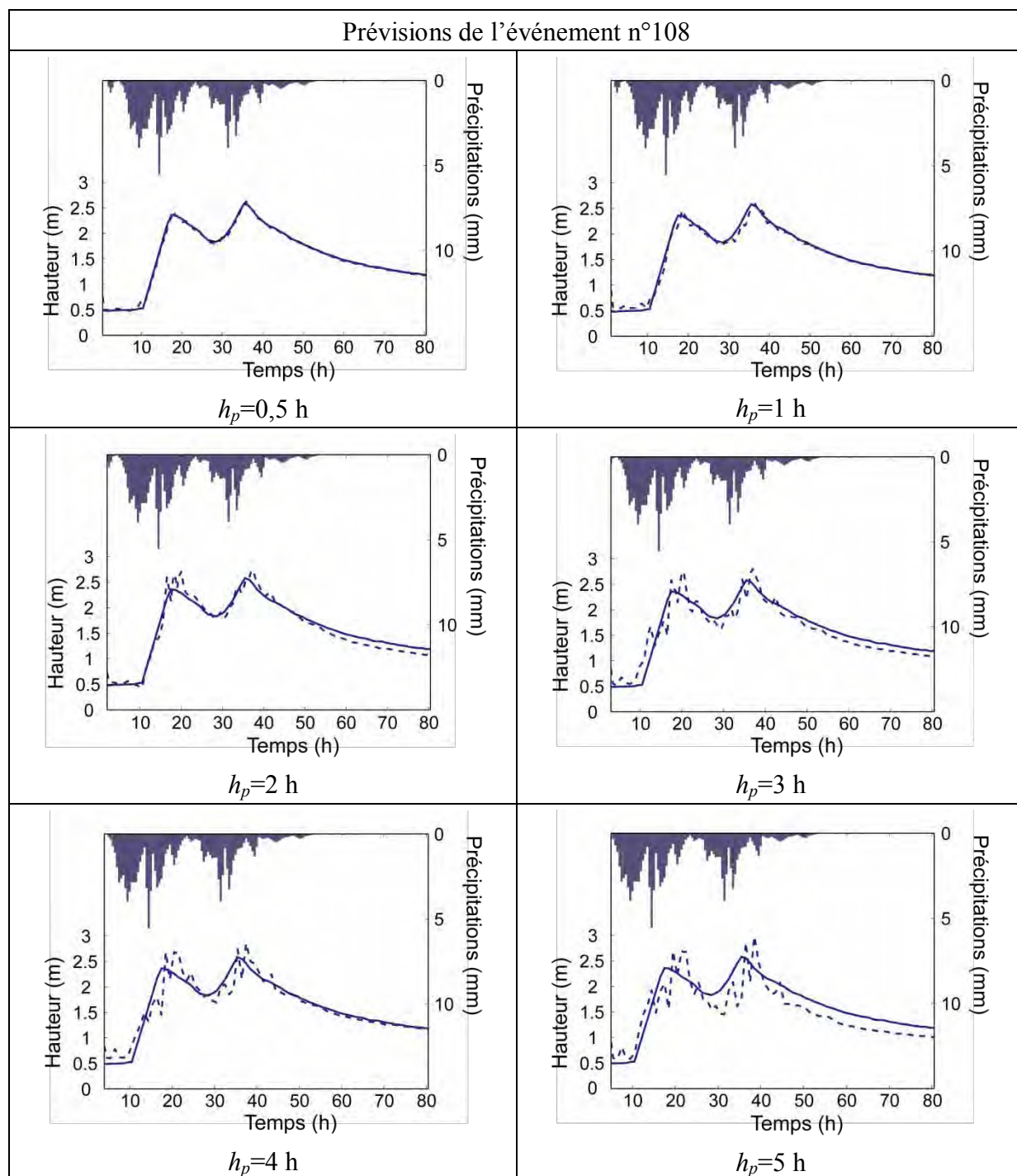


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,72	0,57	0,38	0,78	0,67	0,64	0,63
$R^2$ (Critère de Nash)	1,00	0,99	0,93	0,95	0,88	0,82	0,93
Critère de hauteur $C_h$	0,99	0,95	0,89	0,93	0,91	0,84	0,92
$R^2$ (prévision naïve)	0,99	0,97	0,90	0,78	0,64	0,49	0,80
$C_h$ (prévision naïve)	0,97	0,95	0,88	0,80	0,73	0,69	0,84

Critères de qualité de la prévision

## Evénement 108



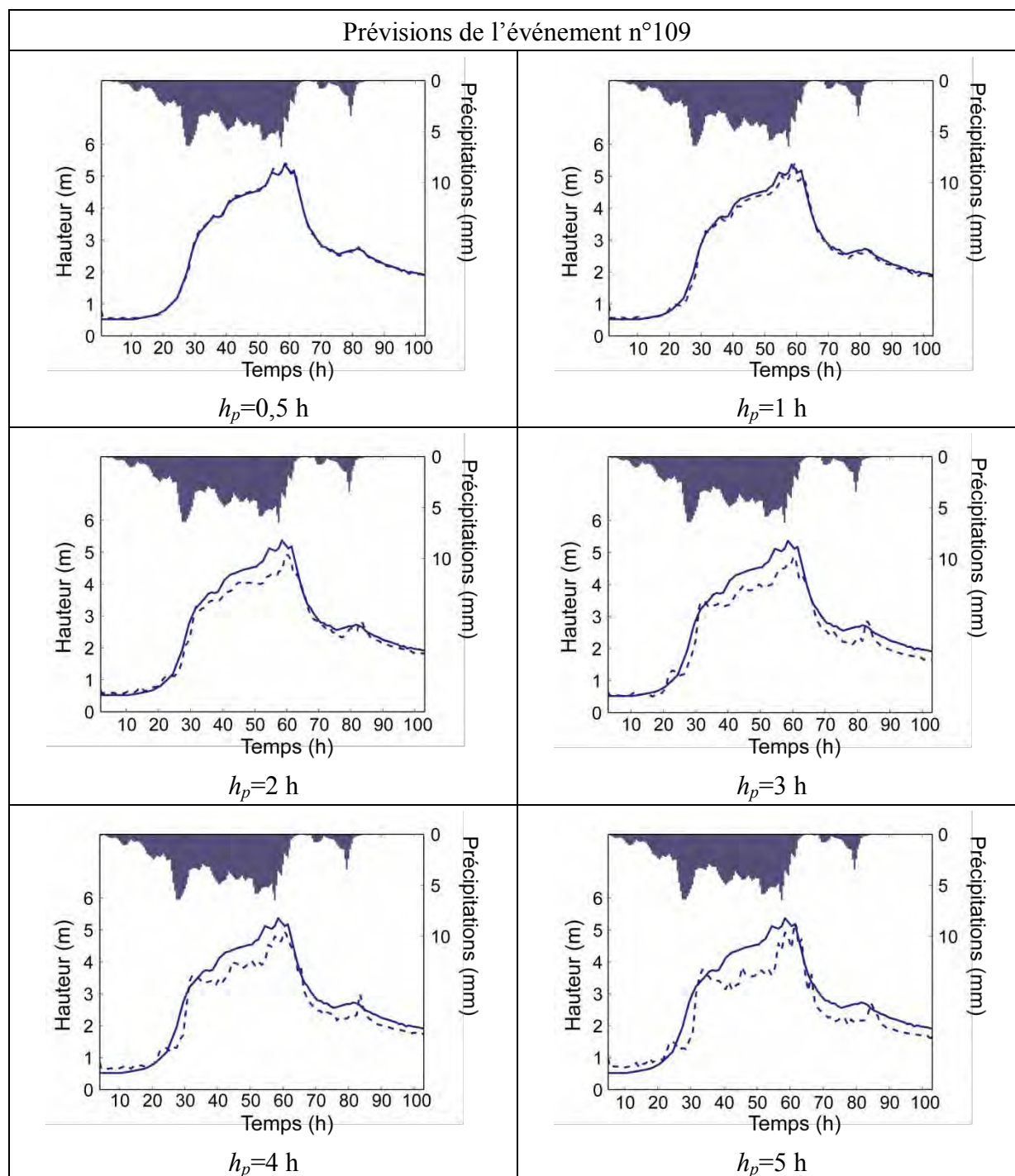
Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,33	0,33	0,57	0,52	0,70	0,50	0,49
$R^2$ (Critère de Nash)	1,00	0,98	0,96	0,90	0,88	0,71	0,91
Critère de hauteur $C_h$	0,98	0,94	0,94	0,90	0,83	0,79	0,90
$R^2$ (prévision naïve)	0,99	0,98	0,90	0,78	0,61	0,41	0,78
$C_h$ (prévision naïve)	0,97	0,93	0,84	0,75	0,67	0,59	0,79

Critères de qualité de la prévision



## Événement 109



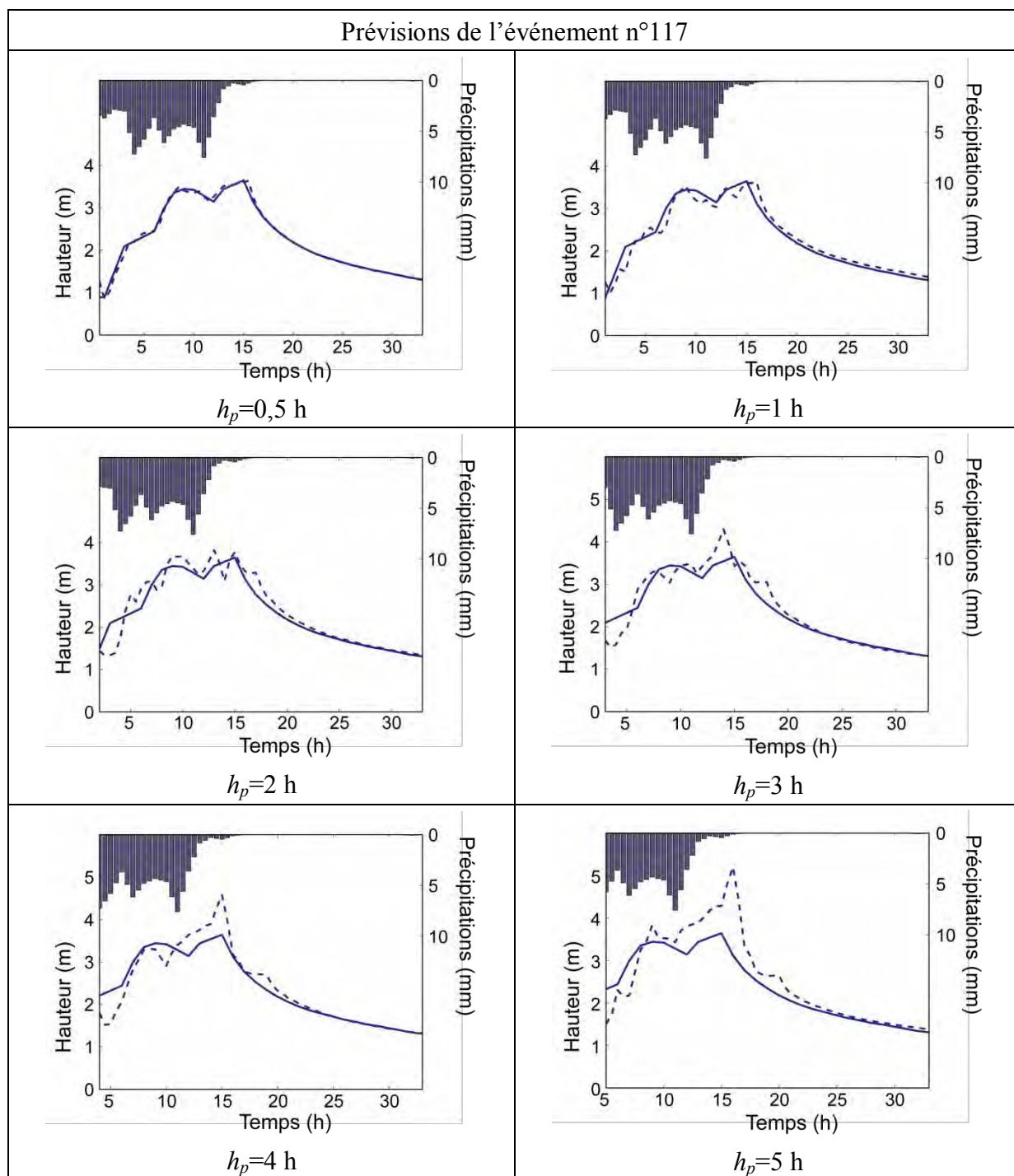
Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.



Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,55	0,11	-0,49	-0,13	0,30	0,12	0,08
$R^2$ (Critère de Nash)	1,00	0,99	0,96	0,93	0,92	0,85	0,94
Critère de hauteur $C_h$	0,99	0,97	0,85	0,88	0,90	0,89	0,91
$R^2$ (prévision naïve)	1,00	0,99	0,97	0,94	0,89	0,83	0,94
$C_h$ (prévision naïve)	0,99	0,98	0,97	0,97	0,96	0,92	0,97

Critères de qualité de la prévision

## Événement 117

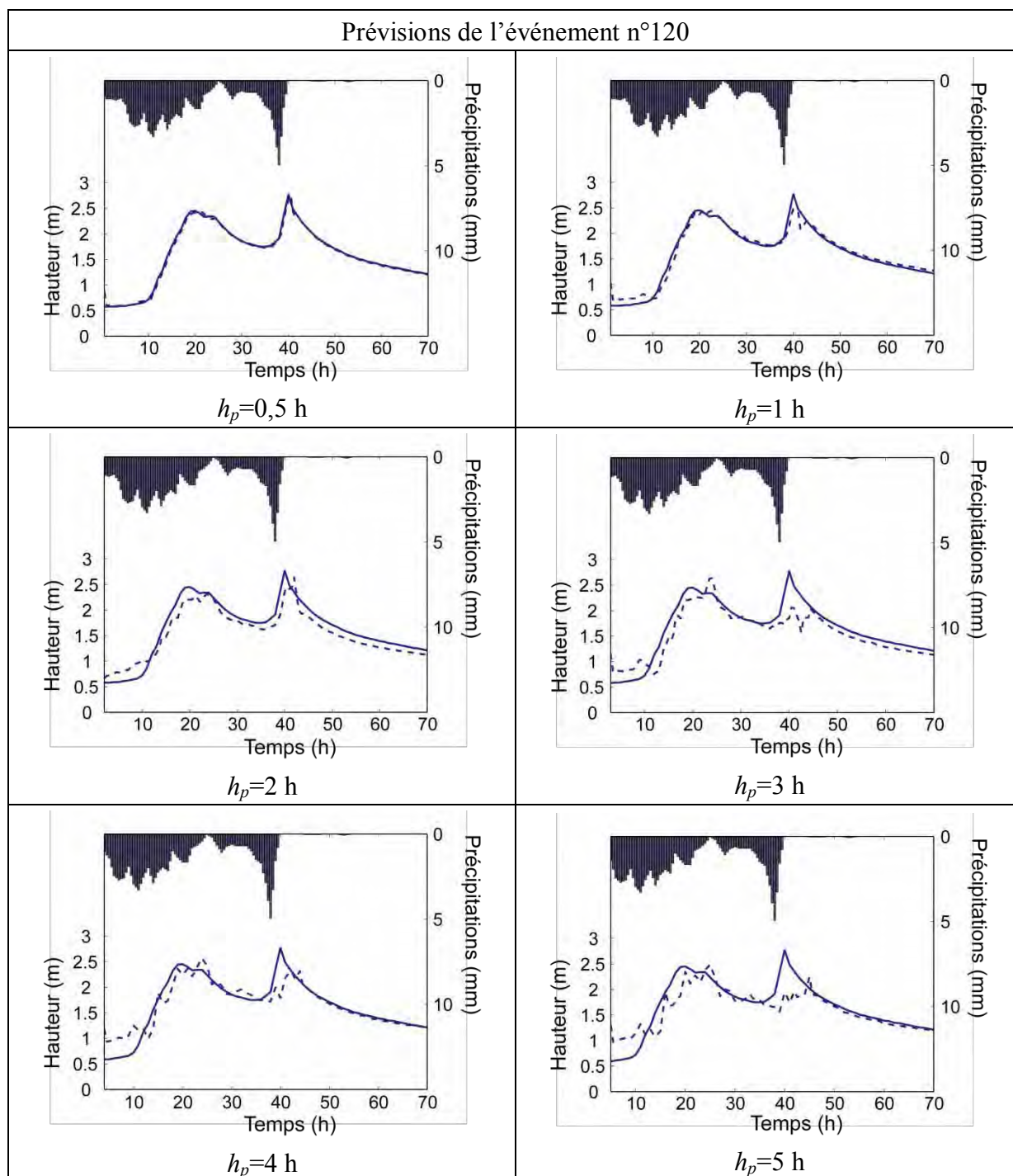


Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,56	0,45	0,67	0,81	0,82	0,61	0,65
$R^2$ (Critère de Nash)	0,99	0,95	0,90	0,89	0,85	0,57	0,86
Critère de hauteur $C_h$	0,98	0,96	0,94	0,92	0,91	0,88	0,93
$R^2$ (prévision naïve)	0,98	0,91	0,68	0,41	0,18	-0,10	0,51
$C_h$ (prévision naïve)	0,97	0,94	0,87	0,82	0,81	0,80	0,87

Critères de qualité de la prévision

## Événement 120



Les précipitations moyennes observées sur le bassin versant (moyenne calculée sur les 6 pluviomètres) sont tracées en ordonnée inversée (échelle de droite) ; la hauteur d'eau à Anduze est tracée en ordonnée positive (échelle de gauche) ; la courbe en trait plein est la hauteur mesurée et la courbe en pointillés la hauteur estimée par le prédicteur.

Horizon de prévision $h_p$ (heures)	0,5 h	1 h	2 h	3 h	4 h	5 h	Moyenne
Critères de Persistance $C_p$	0,46	0,20	0,30	0,30	0,66	0,59	0,42
$R^2$ (Critère de Nash)	0,99	0,97	0,90	0,80	0,83	0,70	0,87
Critère de hauteur $C_h$	0,97	0,94	0,88	0,83	0,86	0,78	0,88
$R^2$ (prévision naïve)	0,90	0,96	0,86	0,71	0,52	0,28	0,71
$C_h$ (prévision naïve)	0,95	0,92	0,84	0,78	0,74	0,70	0,82

Critères de qualité de la prévision

## **Annexe H : Article accepté par la revue « Applied Intelligence »**

### **Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of "Cévenol flash floods"**

Toukourou Mohamed<sup>1</sup>, Johannet Anne<sup>1</sup>, Dreyfus Gérard<sup>2</sup>, Ayral Pierre-Alain<sup>1</sup>

*(1) Ecole des Mines d'Alès, CMGD/LGEI, 6 av. de Clavières, 30319 Alès Cedex, France.*

*(2) ESPCI-Paristech, Laboratoire d'Electronique, 10 rue Vauquelin, 75005 Paris, France.*

**Abstract:** "Cévenol flash floods" are famous in the field of hydrology, because they are archetypical of flash floods that occur in populated areas, thereby causing heavy damages and casualties. As a consequence, their prediction has become a stimulating challenge to designers of mathematical models, whether physics based or machine learning based. Because current, state-of-the-art hydrological models have difficulty performing forecasts in the absence of rainfall previsions, new approaches are necessary. In the present paper, we show that an appropriate model selection methodology, applied to neural network models, provides reliable two-hour ahead flood forecasts.

**Keywords:** *flood forecasting, rainfall-runoff relation, neural networks, dynamic system, generalization.*

### **Introduction**

The prediction of flash floods in populated areas is a major challenge for early flood warning systems. For the watershed under consideration in the present paper, current physics-based models have difficulties performing accurate real-time forecasting in the absence of accurate rainfall forecasts. Therefore, machine-learning models based on past flash flood measurements in the same watershed are attractive alternatives.

After a presentation of the “*Cévenol flash floods*” whose prediction is investigated in the present paper, the current physics-based methods are described cursorily, and a short presentation of past attempts at forecasting flash floods with neural networks is provided. In the subsequent section, two candidate models for nonlinear dynamic process forecasting are presented: recurrent neural networks and feedforward neural networks with time delays; it is shown that, given the specific features of flash flood forecasting in the absence of rainfall predictions, the latter models are more appropriate. The contents and structure of the database is described, and a useful typology of the flood events present therein is provided. The subsequent section describes the model design methodology, with emphasis on variable and model selection by leave-one-sequence-out cross-validation, training and regularization, and independent testing. In the final section, the results are described, and we show that satisfactory two-hour ahead predictions are feasible, thereby opening the way to issuing reliable population warnings in real time.

## Flood Forecasting

Flood forecasting involves essentially finding an appropriate relation between rainfalls (past and present) and future river runoff. For heterogeneous watersheds, this is a very difficult task, for various reasons. First, the rainfall-runoff relation depends on the season due to evaporation and evapotranspiration (consumption and evaporation of water by vegetation) [1], which may lead to significant spatial variations of the initial water storage [2]. Moreover, the watershed response also depends on soil moisture and land use. As a consequence, the rainfall-runoff relation is nonlinear, and it depends on variables that are difficult to measure and/or to estimate [2], [3], [4].

### Flash Floods of Gardon d’Anduze

The Experimental Digital Watershed (EDW) program<sup>16</sup> was initiated by the French Ministry of Environment (MEEDADT) in order to compare the real-time performances of different forecasting models on a few test watersheds. Three EDW test sites were proposed: the *Meuse* and the *Marne* watersheds for slow floods and “water table” floods respectively, and the *Gardon* watershed at *Anduze* for flash floods. Flash floods are floods that exhibit a sharp flow peak occurring on a very short time scale (less than 4-6 hours in the case of the *Gardon* watershed) after a very intense rainfall. The prediction of the latter flood type is still an open issue, because of its complexity, which is due mainly to the spatial heterogeneity of the rainfall and of the soil moisture [2] [4]. In the absence of accurate measurements of these important variables, it is very difficult for physical models to predict accurately the evolution of the water levels or flows.

In view of the damages caused by flash floods in populated areas (over 100 fatalities over the past two decades in the southeast of France, and 1.2 billion € of

---

<sup>16</sup> French acronym: BVNE for Bassin Versant Numérique Expérimental

economical damages for the sole event of 9<sup>th</sup> September 2002 [7], [8]), predictive methods that do not require the accurate knowledge of the above factors are potentially very useful.

The *Gardon* catchment exhibits the typical behavior that the present article aims at modeling: floods are very irregular and may rise up to several meters in a few hours. In addition, it is vulnerable: a population lives and works in this basin, which explains the huge damage costs and loss of human lives.

The *Gardon d'Anduze* catchment, sub-catchment of the *Gardon* catchment (Rhône river tributary) is located in the southeast of France as shown in Figure 1, in the *Cévennes* mountainous area. The basin area is 546 km<sup>2</sup>, the catchment is mountainous with large mean slopes (40%) and high soil infiltration capacity, which explains the velocity of the floods. The basin contains three main geological units: schists (60 %), granite (30%) and limestone (10%), which accounts for the heterogeneity of soil moisture and permeability.

The *Anduze* catchment is subjected to very intense storms delivering huge amounts of water: for example, a 600 mm rainfall was recorded in the *Anduze* rain gauge in less than 12 hours in 2002<sup>17</sup>. These storms occur most frequently during autumn when the Mediterranean Sea is almost warm.

Rainfalls and flows were measured both for very intense and for “moderate” flood events, over a period of thirteen years. Spatially distributed rainfall measurements were provided by six rain gauges. The list of recorded flash flood events for the *Gardon* watershed at *Anduze* is reported in Table 1.

---

<sup>17</sup> For comparison purposes, 600 mm is approximately the mean yearly rainfall in Paris (France).



N°	Date	Duration (hours)	Cumulated rainfall (cm)	Maximum discharge (m <sup>3</sup> /s)	Maximum water level (m)	Mean Flow (m <sup>3</sup> /s)
1	11/ 21-24/1994	35	18.9	491	3.71	181
3	10/4-5/1995	54	32	975	5.34	175
4	10/13-14/1995	92	19	864	5	108
8	11/10-12/1996	82	16	268	2.71	33
13	9/28-29/2000	46	18	800	4.80	74
19	9/9/2002	29	41	2742	9.71	511
22	9/24-25/2006	23	6	186	2.24	21
23	10/19-20/2006	55	20	1436	6.61	181
24	11/17/2006	34	6	275	2.75	65
25	11/20-23/2007	70	12	264	2.69	139
107	11/5-7/1997	74	27	624	4.20	225
108	11/26-27/1997	66	11	244	2.58	117
109	12/18-19/1997	104	38	985	5.37	327
117	10/20-21/1999	34	12	473	3.64	212
120	11/12-14/2000	71	13	279	2.77	110

Table 1. List of available events between 1994 and 2007 on the *Gardon d'Anduze* catchment (from EDW database). Events are identified by a number, and each row of the table reports the date of the event, its duration, the cumulated rainfalls over the duration of the event (averaged over the 6 rain gauges), the maximum peak discharge (averaged over half an hour), the maximum water level at the *Anduze* gauge station and the mean discharge over the duration of the event. The most intense event is number 19, whose prediction is one of the objectives of the present work.

Usually the flow is not measured directly: it is derived from the water level by means of a “rating curve”. The rating curve must be derived carefully; it depends on the section of the river, on the roughness of the waterbed, on its slope, and it may vary in time due to local changes in the waterbed during floods. In the present study, water levels were forecasted and subsequently converted to discharges using the rating curve. The discharge is the physical value characterizing the river state, while the level is location-specific.

As an additional difficulty, various sensor failures may affect the accuracy of the measurements.

At present, measurements of the *Gardon d'Anduze* catchment are sampled with a five-minute sampling rate and transmitted to the local Flood Forecasting Service which analyzes the incoming data.

## Hydrological Methods for Flood Forecasting

One can distinguish between *simulations*, where the flow at time  $t$  is computed from rainfalls at the same time  $t$  or earlier, and *forecasts*, where the flow is computed ahead of time. Since water flow is due solely to rainfalls, flow forecasting requires either rain forecasts, or *ad hoc* assumptions on future rains, such as null or constant rainfall. The last two assumptions are not satisfactory in the case of flash flood as investigated here, because rainfalls are highly variable in time. Furthermore, these assumptions cannot take into account the spatial variability of rainfalls. Rainfall predictions in the area under investigation are not accurate enough, because the rain generation process itself is not yet clearly understood: it is still a difficult task to forecast accurately when or where the storm will be stabilized, even for recent meteorological models (such as AROME [5]), due to the “backward regenerative system” that feeds the storm with the water of the Mediterranean sea [6]. Hydrological models may belong to one of three categories: (i) physical models, (ii) conceptual models and (iii) statistical models. Some models implement a combination of these approaches. The first two families of models will be presented briefly, before describing neural network based statistical models in detail.

### *Physical modeling*

Physical modeling aims at modeling the physical phenomena that result in flash floods. In the case of the catchment under investigation, the heterogeneity of rainfalls and the behavior of the rainfall-soil couple are the main factors that account for the phenomena: physics-based models must take them into account. For example, the model that is used in the “Experimental Digital Watershed” project, and applied to the *Anduze* catchment [2], estimates, at each time step (hourly), the spatial distribution of the soil water content (with three parameters). In a subsequent step, for six appropriately defined hydrological meshes, two elements of the global discharge are estimated at the pixel resolution of the “Digital Terrain Model”: the subsurface lateral flow (water flowing underground laterally in the few first meters of soil and drained by the river network) and the saturated area runoff (water that flows on the surface because the soil is already saturated due to previous rainfall, thus unable to absorb additional water). Then a geomorphological approach allows an estimation of the time necessary for underground or surface water to reach the river at any location. Six sub-catchments are thus defined, where discharge measurements are performed (one measurement station per catchment), three physical parameters are required to compute the total runoff for each sub-catchment, so that finally eighteen parameters are defined; two additional parameters, accounting for evapotranspiration and for water transfer, are required. The accurate estimation of those physical quantities is a major difficulty of that approach, because the inverse problem does not have a unique solution.

## Conceptual modeling

Conceptual modeling assumes that the rainfall-runoff relation may be described by global concepts, such as the reservoir concept. This approach has been applied to the *Gardon d'Anduze* catchment using the SCS model (Soil Conservation Service) [3]. It is considered that rainfalls fill a “soil reservoir” whose level is computed using a balance between infiltration and discharge. The reservoir is modeled as a first-order linear filter. Therefore, two parameters only are required: the capacity of the reservoir and its time constant. The soil reservoir depends also on the soil characteristics and is thus spatially distributed. The results of the simulations performed with the model depend strongly on the postulated initial level of the soil reservoir.

For all hydrological models addressing flash floods, the prominent variable is the spatially distributed rainfall. In the case of TOPMODEL, variable ranking shows [2] the major relevance of i) distributed rainfalls, and ii) soil moistures. This expert knowledge will be useful for the design of our statistical models, as shown in the “Model design” section.

Moreover, in both conceptual and physical models, rainfall information must be available; therefore, it is very difficult to forecast the future water level in the absence of rainfall forecast. Neural networks, which can learn functions based on rainfall and runoff time series, are thus good candidates for forecasting runoff at a predefined horizon.

## Neural Networks for Flash Flood Forecasting

Although neural networks have been widely publicized for flood forecasting, the prediction of flash floods, in the absence of rainfall predictions, has been seldom investigated by machine learning techniques. In [9], flash flood forecasting is performed by neural networks, radial basis functions networks and nearest-neighbors, in the framework of nonlinear, autoregressive models, taking into account past rainfall measurements, and past runoff measurements; the authors note that “all investigated MLP networks perform significantly worse when applied to the test data sets”. This overfitting phenomenon can probably be traced to the fact that no regularization mechanism was implemented during training. By contrast, early stopping is used for regularization in [10] and [11], but the scope of that research is long-term predictions. In [12] [13], flash floods are predicted by various methods including neural networks. The reported results are satisfactory probably due to the very small size of the basins (2 and 20 km<sup>2</sup>) which simplify the rainfall runoff relation; moreover they are not obtained on an independent test set: therefore, the results are likely to be unduly optimistic. In [14], even though generalization is assessed on an independent test set, the most intense event of the database is always included in the training set, so that the ability of the model to forecast reliably a more intense event is not demonstrated. The importance of such issues as complexity control, variable selection, independent testing, are not always acknowledged when machine learning techniques are used in hydrology. Some well-known techniques are summarized in [15]. In addition, many publications in the field are plagued by various problems (inefficient optimization

techniques, overparameterized models, inappropriate performance indices, and failure to realize that neural networks are just another nonlinear regression technique); as a consequence, results on difficult problems such as the forecasting of extreme floods are often disappointing, although these are events that most urgently need fast, accurate predictions.

In the present work, we show that a rigorous model design procedure, involving model complexity control, variable selection, regularization, and independent testing, leads to satisfactory predictions of extreme events.

## Problem statement and data processing

### Neural network models for flood forecasting

Time plays a functional role in the rainfall-runoff relation: physically, the flow at time  $t$  depends on previous rainfalls. Therefore, discrete-time dynamic models must be designed. In the following, we denote by  $T$  the sampling period of the physical quantity of interest  $y^p(kT)$ , hereinafter abbreviated as  $y^p(k)$ .

Denoting by  $y(k)$  the predicted value of the quantity of interest at discrete time  $kT$ , by  $\mathbf{u}(k)$  the vector of exogenous variables at time  $kT$ , and by  $g_{NN}$  the nonlinear function implemented by a feedforward neural network, the following simple input-output “neural” predictors can be designed [16]:

(i) recurrent neural networks

$$y^p(k) = g_{NN}(y^p(k-1), y^p(k-2), \dots, y^p(k-n), \mathbf{u}(k-1), \mathbf{u}(k-2), \dots, \mathbf{u}(k-m))$$

where  $n$  and  $m$  are positive integers; recurrent neural networks predict the quantity of interest at time  $k$  from its past *predicted* values and from past measured values of the exogenous variables,

(ii) feedforward neural networks with time delays

$$y^p(k) = g_{NN}(y^p(k-1), y^p(k-2), \dots, y^p(k-n), \mathbf{u}(k-1), \mathbf{u}(k-2), \dots, \mathbf{u}(k-m)),$$

which predict the quantity of interest at time  $k$  from its past *measured* values and from past values of the exogenous variables.

It has been shown [16] that the first category of predictors is optimal if the modeled process is subjected to output noise (typically measurement noise), while the second category is optimal if the modeled process is subjected to state noise (typically unknown disturbances acting on the process itself). If both types of noise are present, recurrent models fed both with past predicted variables and past measured variables must be designed.

In the present study, the quantity of interest  $y^p$  is the water flow (or equivalently the water level), and the exogenous variables are the rainfalls and possibly additional factors as described below.

In addition, the purpose here is to forecast, at time  $kT$ , the water flow or water level at time  $(k + f)T$ , where  $f$  is a given integer. The prediction horizon is typically a few hours, while the sampling period is typically a few minutes. In

order to perform such a forecast, a recurrent network might be used by running the predictor  $f$  times at time  $kT$ : that would require the availability of  $\mathbf{u}(k+1)$ ,  $\mathbf{u}(k+2)$ , ...,  $\mathbf{u}(k+f)$ , i.e. the availability of rainfall forecasts on that horizon. This requirement precludes the use of recurrent predictors.

Therefore, the predictors designed in the present study are of the form

$$y(k+f) = g_{NN}^f(y^p(k), y^p(k-1), \dots, y^p(k-n), \mathbf{u}(k), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m))$$

where  $g_{NN}^f$  is implemented as a feedforward neural network whose number of variables and number of hidden neurons depend on the forecasting horizon  $f$ .

Therefore, given a forecasting horizon  $f$ , one has to find the appropriate time window  $m$  for rainfalls, the appropriate time window  $n$  for past water levels, the relevant exogenous variables, together with the appropriate complexity in terms of number of hidden neurons, given the available data. We perform this task as described below in the “model and variable selection” subsection of the “model design” section.

## Data acquisition

Rainfalls are measured by rain gauges, which are the most accurate sensors available at present. However, they provide local information, so that the heterogeneity of rainfalls is a major problem for flash floods. For instance, in the case of event 19, which is of particular interest in the present study, the cumulated rainfall was three times as large in *Anduze* as in *Soudorgues*, which is only fifteen kilometers away. In order to obtain a more global picture, radar acquisitions of rainfalls with a definition of  $1 \text{ km}^2$  have been performed since 2002, but the number of flood events thus monitored in the catchment under investigation is still too small for reliable use in a machine learning approach.

The sampling period for water level measurements during flood events was 1 h before 2002 and has been 5 minutes since 2002. In the present work, a sampling period of 30 mn was chosen, which is appropriate considering the 2-4 hours rise time of this catchment. For events that occurred before 2002, re-sampling was performed by linear interpolation.

## Database structure

The water level of the catchment under investigation is sampled all year long with a sampling period of one hour, but the sampling period is shortened to 5mn when flash floods occur. Therefore, the database was organized as follows: at the end of each flood event, a non-flood (zero or weak rainfall) time period of 15 hours (more than twice the concentration time of the basin, i.e. the time that is necessary for water to flow from the most distal point of the basin to the catchment where the measurement is performed) was kept in the database after each event. The flood events were subsequently concatenated into a single time series (Figure 2). Between two concatenated events, the non-flood zone is used simultaneously as rainfall for the end of the present event and as the rainfalls for the future event. A transition zone is also introduced between two adjacent events in order to decrease

the outflow discontinuity between the end of an event (high discharge) and the beginning of the adjacent event (low discharge). The resulting database contains 1743 measurements of rainfalls and water levels at *Anduze*.

## Typology of flood events

In order to perform (i) training, (ii) model and variable selection, and (iii) testing, three data sets are necessary: a training set, a validation set and a test set. They must be chosen appropriately. To this end, a typology of the available events was created as follows.

The 15 events listed in Table 1 were clustered manually into three categories depending on their intensity, time length and number of peaks: an event was classified as “very intense” if the maximum water level exceeded 4.5 m; it was classified as “long” if its duration exceeded 50 h; finally, single-peak events were separated from events with multiple water level peaks. This resulted in 8 categories, which are listed in Table 2.

Very intense	Long	Single-peak	Class and number of elements
Y	Y	Y	A (3)
Y	Y	N	B (1)
Y	N	Y	C (2)
Y	N	N	D (0)
N	Y	Y	E (0)
N	Y	N	F (5)
N	N	Y	G (3)
N	N	N	H (1)

Table 2. List of classes and of their number of elements

As might be expected, classes D and E are empty: an intense and short event cannot have more than one peak, and intense but long events necessarily have several peaks.

## Model design

As indicated in the previous section, the constraints of the problem lead naturally to choosing a model of the form

$$y_{k+f} = g_{NN}^f(y^p_k, y^p_{k-1}, \dots, y^p_{k-n}, \mathbf{u}_k, \mathbf{u}_{k-1}, \dots, \mathbf{u}_{k-m})$$

where  $y$  is the estimated water level,  $y^p$  is the measured water level,  $\mathbf{u}$  is the vector of exogenous variables  $g_{NN}^f$  is the function implemented by a feedforward neural network with one layer of hidden neurons with tanh nonlinearities, and a linear output neuron.

As pointed out in section “hydrological method for flash flood modeling”, the vector of exogenous candidate variables includes the rainfalls from 6 rain gauges (spatially distributed rainfalls), and the soil moisture (Soil Water Index, given by the ISBA model [17]). The actual relevance of the candidate variables must be assessed by a variable selection procedure.  $n$  and  $m$  are positive integers that must also be chosen by that procedure. The latter is described below, in the section entitled “model and variable selection”.

## **Training**

The usual least squares cost function was optimized by the Levenberg-Marquardt algorithm; the gradient of the cost function was computed by backpropagation.

## **Regularization**

Regularization was performed by early stopping on event 13, which appears to be prototypical of the available events. For more details on training and regularization, see e.g. [18].

## **Model and variable selection**

Model and variable selection aim at finding the appropriate number of hidden neurons, the appropriate values of  $m$  and  $n$ , and the relevance of the soil moisture variable. The procedure was a “leave-one-sequence-out” procedure, inspired by the usual leave-one-out cross-validation method. First, sequence 13, which is used for early stopping (see section “Regularization”), and sequence 19, which is used for testing (see section “Test”) are set apart. From the set of the remaining 13 sequences, each sequence in turn is extracted, a predictor is trained on the other 12 sequences, and its performance on that sequence is computed. The leave-one-sequence-out score is the average of the sum of squared prediction errors, over all left-out sequences. However, since the accuracy of the predictions of intense sequences is more critical than the accuracy of the predictions of moderately intense ones, the following variant was found more successful: instead of leaving out each sequence in turn, the least intense nine sequences were always kept in the training set, and one sequence among the most intense four sequences (3, 4, 109 and 23) was left out in turn.

At the end of the procedure, the combination of  $n$ ,  $m$ , of the number of hidden neurons, and of the presence or absence of the soil moisture variable, that provides the best leave-one-sequence-out score was selected. Such a model was subsequently trained with all available sequences except the early stopping sequence and the test sequence.

## **Test**

The quality of the resulting model was assessed on the test sequence, *i.e.* event 19 (class C), whose intensity made history in 2002. Several quality criteria can be considered.

The coefficient of determination of the regression, known to hydrologists as Nash's criterion, is defined as

$$R^2 = 1 - \frac{\sum_{k \in \text{test sequence}} (y^p_k - y_k)^2}{\sum_{k \in \text{test sequence}} (y^p_k - \overline{y^p})^2}$$

where  $\overline{y^p}$  is the mean of the observations.  $R^2 = 1$  if the model predicts the test data with perfect accuracy, and  $R^2 = 0$  if the model simply predicts the mean of the observations.

For forecasting, the persistence coefficient  $C_p$  is of special interest. For a forecasting horizon  $f$ , it is defined as [19]:

$$C_p = 1 - \frac{\sum_{k \in \text{test set}} (y^p_{k+f} - y_{k+f})^2}{\sum_{k \in \text{test set}} (y^p_k - y^p_{k+f})^2}$$

$C_p$  ranges from  $-\infty$  to 1.  $C_p = 1$  if the model predicts the test data with perfect accuracy, and  $C_p = 0$  if the predictor is perfectly dumb, *i.e.* if it predicts that the future value of the quantity of interest will be equal to the present one. A predictor with negative  $C_p$  performs worse than a dumb predictor.

## Results

In order to obtain baseline results, static predictors of the form

$$y_{k+f} = g_{NN}^f(\mathbf{u}_k, \mathbf{u}_{k-1}, \dots, \mathbf{u}_{k-m}),$$

where  $\mathbf{u}(k)$  is the vector of exogenous variables, will first be described. Dynamic predictors will be considered in the subsequent section.

### Static predictors

The procedure described in section “Model and variable selection” showed that the best leave-one-sequence-out scores were obtained by networks having 2 hidden neurons, with  $m = 11$  (12 rainfall values, sampled at 30-min intervals) for thirty-minute ahead forecasts,  $m = 7$  for one-hour ahead forecasts, and  $m = 3$  for two-hour ahead forecasts. Soil moisture was not found to be relevant: it was not included in vector  $\mathbf{u}$ . An additional input variable was found to be relevant, and was therefore included in vector  $\mathbf{u}$ : the rainfall accumulation from the beginning of the event.

For two forecasting horizons out of three, the persistence coefficient was negative, showing that the predictor is less efficient than a dumb predictor. Figure 3 shows the two-hour ahead forecasts of event 19 and Table 3 the values of the quality criteria.



Forecasting horizon	thirty-minutes	one hour	two hours
$R^2$ (Nash criterion)	0.83	0.79	0.76
$C_p$	-2.68	-0.32	0.45

Table 3. Quality criteria for static network (event 19).

### Dynamic predictors: feedforward neural networks with time delays

In the present section, we consider predictors of the form

$$y_{k+f} = g_{NN}^f(y_{k-n}^p, y_{k-n+1}^p, \dots, y_k^p, u_{k-m}, u_{k-m+1}, \dots, u_k)$$

As in the case of static predictors, the procedure described in section “Model and variable selection” resulted in the selection of  $m$  and of the number of hidden neurons, and in the assessment of the relevance of the candidate exogenous variables. In addition, the appropriate value of  $n$  was selected by the procedure.

The results are summarized as follows:

- $n = 1$  for all values of  $f$  (present value and one past value of water level),
- 2 hidden neurons for thirty-minute and one-hour ahead forecasts, 5 hidden neurons for two-hour ahead forecasts,
- $m = 4$  for thirty-minute ahead forecasts,  $m = 5$  for one-hour and two-hour ahead forecasts.
- soil moisture and cumulated rainfalls from the beginning of the event were found irrelevant.

These predictors are much more efficient than the static predictors. Table 4 summarizes the values of the coefficient of determination  $R^2$  and of the persistence coefficient  $C_p$  for various forecasting horizons, for the most intense, single-peak event (19) and for a more moderate, multiple-peak event (107, class F). The results on event 107 were obtained by exactly the same procedure as for event 19, except that event 19 was in the training set and that event 107 was used neither for training nor for validation. In both cases the results are very satisfactory. For event 19, the prediction of the peak is performed two hours before its occurrence, but the rainfall peak occurs two hours before the peak of the discharge; this explains in part why that particular event is very difficult to predict accurately. To the best of our knowledge, no forecasting method has ever achieved a comparable accuracy on such an extremely intense flash flood as event 19.

Event	Forecasting horizon	$R^2$	$C_p$
19	½ h	0.98	0.60
	1 h	0.92	0.51
	2 h	0.86	0.68
107	½ h	1.00	0.72
	1 h	0.99	0.77
	2 h	0.98	0.76

Table 4. Summary of test results obtained on two different events at three different forecasting horizons.

Figure 4-a, -b, -c, and Figure 5-a, -b, -c show the observed and forecast hydrographs for the above events.

### **Assessment of the relevance of the predictor with respect to present vigilance thresholds**

From the end user's point of view, it is important to assess the behavior of the predictor in the framework of the present flood warning system. Four vigilance levels, defined by the SCHAPI, are currently in use:

- Red: major risk of high water level with serious threats to people and property,
- Orange: risk of high water level with considerable overflow liable to have a significant effect on daily life and on the safety of people and property,
- Yellow: risk of high or rapidly rising water not involving significant damage but requiring particular vigilance in the case of seasonal and/or outside activities,
- Green: no particular vigilance required.

For the *Gardon* at *Anduze*, the thresholds are defined as follows by the local Flood Forecasting Service,

- discharge < 800 m<sup>3</sup>/s: green
- 800 m<sup>3</sup>/s < discharge < 1600 m<sup>3</sup>/s: yellow
- 1600 m<sup>3</sup>/s < discharge < 2800 m<sup>3</sup>/s: orange
- discharge > 2800 m<sup>3</sup>/s: red.

Therefore, predicting whether a given vigilance threshold will be reached may be viewed as a classification problem. This task is much less demanding than the prediction problem addressed in the present paper: the classification may be correct although the predicted water level, or the timing of the peaks, may be quite erroneous.

For the different prediction horizons  $f$ , the predicted and the observed vigilance levels can be compared (*a posteriori* because the flood warning system was not available in 2002), so that the ability of the model to forecast the risk of the event, in the framework of the present warning system, can be assessed. Table 5 presents the results of this analysis. The agreement between the vigilance forecasts and the issued vigilance warning is very good: for 10 events out of 14, the risk is correctly classified for all prediction horizons. When misclassifications occur, it is only for high forecasting horizons ( $\geq 3$  h), with two false negatives. The 2002 event (event 19, with a discharge peak value of 2742 m<sup>3</sup>/s)) is classified in orange vigilance but it is very close to the red level.

n° event	Observed	$f=30$ mn	$f=1$ h	$f=2$ h	$f=3$ h	$f=4$ h	$f=5$ h
	Vigilance level	Vigilance level	Vigilance level	Vigilance level	Vigilance level	Vigilance level	Vigilance level
01	Green	Green	Green	Green	<b>Yellow</b>	Green	Green
03	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
04	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
08	Green	Green	Green	Green	Green	Green	Green
19	Orange	Orange	Orange	Orange	Orange	Orange	<b>Yellow</b>
22	Green	Green	Green	Green	Green	Green	Green
23	Yellow	Yellow	Yellow	Yellow	<b>Green</b>	<b>Green</b>	<b>Green</b>
24	Green	Green	Green	Green	Green	Green	Green
25	Green	Green	Green	Green	Green	Green	Green
107	Green	Green	Green	Green	Green	Green	Green
108	Green	Green	Green	Green	Green	Green	Green
109	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
117	Green	Green	Green	Green	Green	Green	<b>Yellow</b>
120	Green	Green	Green	Green	Green	Green	Green

Table 5. Forecasted vigilance level for all events and all forecasting horizons. Boldface characters highlight misclassifications.

## Conclusion

Forecasting flash floods in populated areas is an important and difficult task, which is traditionally approached by physics-based models or by hydrology-based conceptual models. In the present paper, we have shown that such forecasts can be made with very satisfactory accuracy by machine-learning methods, in the absence of rainfall predictions, provided an appropriate methodology is used. Leave-one-sequence-out cross-validation was used for model and variable selection in the framework of feedforward models with time delays. Regularization was performed by early stopping. Blind testing was performed on very intense events, which caused large damages. The accuracy of the forecasts, for the water level as well as for the vigilance level, opens the way to early warnings of the population in the area under investigation. Future research will involve the application of the methodology to other flash-flood areas in France. In

addition, support vector regression and dynamic modeling with kernel methods [20], on the same data, will be investigated.

## **Acknowledgements**

This work was funded in the framework of Experimental Digital Watershed (BVNE), a program initiated by the SCHAPI of the French Ministry of Environment (MEEDADT).

We are very grateful to MM. Pierre Roussel-Ragot, Georges-Marie Saulnier, Bernard Vayssade and Marc Vinches for helpful discussions.

We also wish to thank M. Dominique Bertin for his highly effective collaboration in the design and implementation of the Neural Network simulation tool: RnfPro.

## **Literature references**

- 1\_Oudin L, Hervieu F, Michel C, Perrin C, Andréassian V, Anctil F, Loumagne C (2005) Which potential evapotranspiration input for a lumped rainfall-runoff model ? Part 2- Toward a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of hydrology* 303:290-306.
- 2\_Le Lay M, Saulnier G-M (2007) Exploring the Signature of Climate and Landscape Spatial Variabilities in Flash Flood Events: Case of the 8–9 September 2002 Cévennes-Vivarais Catastrophic Event”. *Geophysical Research Letters*, 34: L13401.
- 3\_Bouvier C, Ayral P-A, Brunet P, Crespy A, Marchandise A, Martin C (2006) Recent Advances in Rainfall-Runoff Modelling: Extrapolation to Extreme Floods in Southern France. *International Workshop on Hydrological Extremes. Observing and modelling exceptional floods and rainfalls Rende (CS)*, 3-4 May.
- 4\_Estupina-Borrell V, Chorda J, Dartus D (2005) Flash-Flood anticipation”. *CR Geosciences* 337:1109-1119.
- 5\_Bouttier, F. The Arome mesoscale project. *Proceedings of the ECMWF*, September 2003
- 6\_Delrieu G, Ducrocq V, Gaume E, Nicol J, Payrastré O, Yates E, Kirstetter P, Andrieu H, Ayral P, Bouvier C, Creutin J, Livet M, Anquetin S, Lang M, Neppel L, Obled C, Parent-du-Châtelet J, Saulnier G-M, Walpersdorf A, Wobrock W (2004) The catastrophic flash-flood of 8-9 september 2002 in the Gard region, France: A first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory. *Journal of Hydrometeorology*, 6:34-52.

- 7\_Gaume E, Livet M, Desbordes M, Villeneuve J-P (2004) Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999. *Journal of Hydrology*, 286:135-154.
- 8\_Sauvagnargues-Lesage S, Ayrat P-A (2007) – Using GIS for emergency management – A case study during the 2002 and 2003 South East flooding in France – *International Journal of Emergency Management*, 4, 4:682-703.
- 9\_Piotrowski A, Napiorkowski J-J, Rowinski P-M (2006) Flash Flood Forecasting by Means of Neural Networks and Nearest Neighbour Approach – A Comparative Study. *Nonlin. Processes Geophys.* 13: 443-48.
- 10\_Coulibaly P, Anctil F, Rasmussen P, Bobée B (2000) A recurrent Neural Networks approach using indices of low frequency climatic variability to forecast regional annual runoff. *Hydrological Processes* 14:2755-2777.
- 11\_Coulibaly P, Anctil F, Bobée B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230:244-257.
- 12\_Sahoo G-B, Ray C (2006) Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology* 317:63-80,
- 13\_Sahoo G.B, Ray C, de Carlo E-H (2006) Use of neural networks to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *Journal of Hydrology* 327:525-538.
- 14\_Schmitz G-H, Cullmann J (2008) PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments. *Journal of Hydrology* 360:1-14.
- 15\_Giustolisi O, Laucelli D (2005) Improving Generalization of Artificial Neural Networks in Rainfall-Runoff Modelling. *Hydrological Sciences Journal* 50:439-457.
- 16\_Nerrand O, Urbani D, Roussel-Ragot P, Personnaz L, Dreyfus G (1994) Training Recurrent Neural Networks : Why and How ? An Illustration in Process Modeling. *IEEE Transactions on Neural Networks* 5:178-184.
- 17\_Noilhan J, Mahfouf J-F (1996) The ISBA land surface parameterisation scheme. *Global and Planetary Change* 13:145-159.
- 18\_Dreyfus G (2005) *Neural networks, methodology and applications*. Springer

19\_Kitadinis P-K, Bras R-L (1980) Real-time forecasting with a conceptual hydrologic model: 2 applications and results. *Water Resour. Res.* 16:1034–1044.

20\_Qu H-N, Oussar Y, Dreyfus G, Xu W (2009) Regularized Recurrent Least Squares Support Vector Machines. International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, Shanghai.

## Figure captions

Figure 1. The *Gardon d'Anduze* watershed and its location in France. Discharges are measured in *Anduze* and cumulated rainfalls are measured by six rain gauges distributed over the watershed in order to obtain information about the spatial variability of rain

Figure 2. Concatenation of events 1, 3 and 4. Solid line: water level; dashed line: rainfall averaged over the six rain gauges

Figure 3. Limnograph of event 19. Solid line: observation; dotted line: two-hour ahead forecast by a static network.  $R^2$  and  $C_p$  are 0.76 and -0.45 respectively

Figure 4-a. Limnograph of event 19. Solid line: observation; dotted line: thirty-minute ahead forecast

Figure 4-b. Limnograph of event 19. Solid line: observation; dotted line: one-hour ahead forecast

Figure 4-c. Limnograph of event 19. Solid line: observation; dotted line: two-hour ahead forecast

Figure 5-a. Limnograph of event 107. Solid line: observation; dotted line: thirty-minute ahead forecast

Figure 5-b. Limnograph of event 107. Solid line: observation; dotted line: one-hour ahead forecast

Figure 5-c. Limnograph of event 107. Solid line: observation; dotted line: two-hour ahead forecast

## Figures

Figure 1

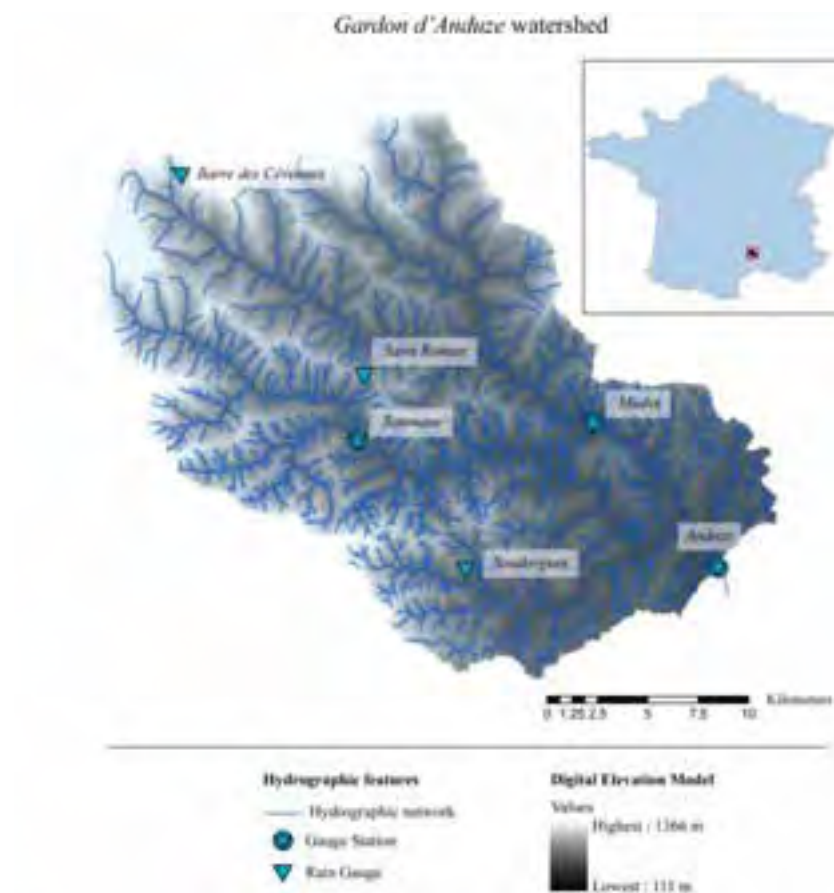


Figure 2

Figure 3

Figure 4

Figure 5







## Application de l'apprentissage artificiel à la prévision des crues éclair

### Résumé :

L'objet de cette thèse est d'appliquer l'apprentissage statistique à la prévision des crues éclair cévenoles qui ont occasionné des pertes humaines et des dégâts considérables durant ces 20 dernières années. Les travaux s'inscrivent dans le cadre du projet *Bassin Versant Numérique Expérimental Gardons*. Dans ce contexte, ce mémoire se propose de présenter les travaux effectués pour réaliser la prévision des crues du Gardon d'Anduze à Anduze, jusqu'à un horizon de prévision de l'ordre de grandeur du temps de concentration, en l'absence de prévision de pluie, par des méthodes d'apprentissage statistique.

La première partie de ce mémoire s'attache donc à présenter la famille de fonctions choisies pour réaliser cette tâche - les réseaux de neurones utilisés pour la prévision du comportement d'un processus dynamique non linéaire -, les propriétés fondamentales qui justifient leur utilisation (l'approximation universelle et la parcimonie), ainsi que les méthodes connues pour éviter le surajustement.

La deuxième partie de ce mémoire présente le cours d'eau qui constitue l'objet de l'étude : le Gardon d'Anduze et ses Gardonnades ainsi que les études hydrologiques qui y sont consacrées. La troisième partie s'intéresse à la mise en œuvre de méthodes de régularisation connues : la modération des poids et l'arrêt précoce. L'efficacité de ces méthodes est connue, mais il a fallu les adapter dans ce travail pour parvenir à estimer le comportement très intense d'un événement extrême. Le mémoire montre comment éviter les phénomènes de spécialisation observés en fonction de l'ensemble d'arrêt, et il introduit la validation croisée partielle, qui est en fait spécialisée sur les événements intenses. C'est grâce à l'établissement d'une typologie des événements, et grâce à une distribution judicieuse de ceux-ci dans les différents sous-ensembles d'apprentissage, de validation, d'arrêt et de test, qu'une procédure générale a pu être établie pour définir ces sous-ensembles et concevoir le modèle de prévision.

Les performances du modèle obtenu en prévision permettent d'envisager qu'une alerte fiable soit disponible sur Internet par l'intermédiaire du SCHAPI et de son site *vigicrues*, rendant ainsi une aide précieuse aux populations.

Ainsi, ce mémoire établit que, contrairement à ce que l'on peut lire dans de nombreuses publications sur les réseaux de neurones appliqués à la prévision des crues, il n'est pas impossible de généraliser à un événement plus intense que ceux de l'ensemble d'apprentissage et il n'est pas non plus impossible d'éviter le surajustement au bruit dans les mesures.

C'est également pour cela que les perspectives ouvertes par ce travail sont particulièrement importantes, tant pour intégrer de l'adaptativité dans le modèle que pour généraliser leur application à des bassins versants non jaugés.

## Machine Learning to Flash Flood Forecasting

### Abstract:

The need for accurate predictions of flash floods has been highlighted by the recent occurrences of catastrophic floods. The scope of this thesis is therefore to apply machine learning to forecast *cévenol* flash floods, which have caused casualties and huge damage in France over the last 20 years. The work was performed in the framework of the *Bassin Versant Numérique Expérimental Gardon* project, initiated by the French Ministry in charge of Sustainable Development. In this context, this work addresses the capability of machine learning to provide flood forecasts in the absence of rainfall forecasts.

The first part of this manuscript describes the family of functions chosen in the present study - neural networks-, their ability to forecast the behavior of non-linear dynamic processes, their fundamental properties (universal approximation and parsimony), as well as the conventional methods used to prevent overfitting.

The second part of this work presents the river under investigation, the *Gardon d'Anduze*, as well as related hydrological studies.

The third part presents the application of two traditional regularization methods: early stopping and weight decay. In order to allow the prediction of very intense floods, an original variable selection method is proposed: "partial cross validation". After careful variable and model selection, the ability of models, obtained by either regularization method, to predict the most dramatic event of the database (September 2002) is assessed, thereby allowing an early warning of the populations.

Thus, this work demonstrates that, in contrast to statements found in many publications on neural networks applied to flash-flood forecasting, the prediction of an event that is more intense than the events present in the database is feasible, provided a rigorous methodology is used. For this reason, this work opens the way to making current models more adaptive, and to applying the method to ungauged basins.